



Éléments de statistiques

A. Qu'est-ce qu'une variable aléatoire ?

C'est une fonction mesurable définie sur l'ensemble des éventualités, c'est-à-dire l'ensemble des résultats possibles d'une expérience aléatoire. Ce furent les jeux de hasard qui amenèrent à concevoir les variables aléatoires, en associant à une éventualité (résultat du lancer d'un ou plusieurs dés, d'un tirage à pile ou face, d'une roulette, ...) un gain. Un vecteur aléatoire est un vecteur dont ses éléments sont des variables aléatoires.

B. Qu'est-ce qu'une fonction de répartition ?

Définition

La **fonction de répartition**, ou **fonction de distribution cumulative** $F_X(x)$, d'une variable aléatoire discrète X , dans S est un ensemble bien ordonné (où ses éléments sont de manière croissante, p.e. $s_1 \leq s_2 \leq s_3 \leq \dots$ et ses probabilités p_s sont en conséquence, p.e. en posant $p_i = p_{s_i}, i \geq 1$. On a alors, si $x \in [s_i, s_{i+1}[$,

$$F_X(x) = \sum_{1 \leq j \leq i} p_j = \sum_{1 \leq j \leq i} \mathbb{P}(X = j)$$



Éléments de statistiques

C. Qu'est-ce qu'une densité de probabilité ?

Les variables aléatoires réelles est entièrement déterminée par sa « densité de probabilité »

La fonction à intégrer est alors appelée **fonction de densité** ou **densité de probabilité**, égale (ou loi de probabilité) à la dérivée de la fonction de répartition.

Définition

Une fonction **f positive** et intégrable sur R , est une fonction de densité, si est seulement si on a $\forall (a, b) \in R^2$ on aura : $\mathbb{P}(a \leq X \leq b) = \int_a^b f(t)dt$.

Remarque

Toute réalisation x de la variable aléatoire X étant comprise entre $-\infty$ et $+\infty$, on a évidemment $F_x(-\infty) = 0$, $F_x(+\infty) = 1$ et $\int_{-\infty}^{+\infty} f(x)dx = 1$



Éléments de statistiques

D. Qu'est-ce que densité de probabilités conjointes ?

- ✓ Si X et Y deux variables aléatoires **indépendantes**
 - ✓ Si la probabilité de réalisation $f_X(x)$ la variable X est **indépendante** de celle de Y $f_Y(y)$
- ⇒ Alors la densité de probabilités conjointes est comme la suivante :

$$f_{X,Y}(x,y) = f_X(x)f_Y(y)$$

E. Qu'est-ce qu'une probabilité conditionnelle ?

« C'est une **probabilité** d'un **événement** sachant qu'un autre événement a eu lieu. »

Exemple : Si une **carte** d'un jeu est tirée au hasard, on estime qu'il y a une chance sur quatre d'obtenir un cœur ; mais si on aperçoit un reflet rouge sur la table, il y a maintenant une chance sur deux d'obtenir un cœur. Cette seconde estimation correspond à la **probabilité conditionnelle** d'obtenir un cœur sachant que la carte est rouge. On a donc $\mathbb{P}(X|Y) = \frac{\mathbb{P}(X \cap Y)}{\mathbb{P}(Y)}$ Où $\mathbb{P}(Y) \neq 0$ et $f_{X,Y}(x,y) = f_Y(y|x)f_X(x) = f_X(x|y)f_Y(y)$

Remarque

Si les variables sont indépendantes $\Leftrightarrow f_Y(y|x) = f_Y(y)$ et $f_X(x|y) = f_X(x)$

Éléments de statistiques



Théorème de Bayes :

Définition :

Probabilités conditionnelles : Quelle est la probabilité qu'un événement se produise **sachant** qu'un autre événement s'est déjà produit.

Le théorème de Bayes est cette forme



$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

Exemple : Supposons qu'on ait une classe de lycéens. Soit A et B les deux événements suivants :

- L'événement A : l'élève est une fille.
- L'événement B : l'élève pratique l'allemand.

Quelle est la probabilité qu'on choisisse au hasard **une fille pratique l'allemand** ?

Le théorème de Bayes permet de calculer ce genre de probabilité. Comme suit :

Éléments de statistiques



Théorème de Bayes :

Exemple :

Notons P la probabilité d'un événement en se basant sur $\mathbb{P}(X|Y) = \frac{\mathbb{P}(X \cap Y)}{\mathbb{P}(Y)}$ où $\mathbb{P}(Y) \neq 0$

$$P(\text{élève est une fille ET élève pratique allemand}) = P(\text{élève est une fille}) * P(\text{élève pratique allemand} | \text{élève est une fille})$$

Selon Théorème de Bayes, on obtient :

$$P(\text{élève est une fille ET élève pratique allemand}) = P(\text{élève pratique allemand}) * P(\text{élève est une fille} | \text{élève pratique allemand})$$

Ceci est fait grâce à cette formule :

$$P(\text{Allemand} | \text{Fille}) = P(B | A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A|B) * P(B)}{P(A)}$$



Éléments de statistiques

F. Qu'est-ce que l'espérance Mathématique ?

Définition

L'espérance mathématique d'une variable aléatoire Y est :

$$\mathbb{E}(Y) = \int_{-\infty}^{+\infty} y f_Y(y) dy$$

Remarque

L'espérance d'une variable aléatoire constante est égale à cette constante ; par exemple b est une constante, alors $\mathbb{E}(b) = b$

Monotonie : si X et Y sont des variables aléatoires telles que $X \leq Y \Rightarrow \mathbb{E}(X) \leq \mathbb{E}(Y)$

Linéarité : l'espérance est un opérateur linéaire. Pour deux variables aléatoires $\forall X$ et Y (appartiennent au même espace probabilisé) et pour deux nombres réels a et b :

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$$

Produit : en général, l'opérateur espérance ne respecte pas le produit, c'est-à-dire qu'en général $\mathbb{E}(XY) \neq \mathbb{E}(X)\mathbb{E}(Y)$. L'égalité est vraie pour des variables X et Y soient indépendantes.

L'objectif de tout **apprentissage** est d'obtenir une **estimation fiable de l'espérance mathématique** de la grandeur à modéliser. À cet effet, \Rightarrow **il est utile d'introduire le concept d'estimateur**.



Éléments de statistiques

I. Qu'est-ce que Estimateur non biaisé ?

- ❖ un estimateur est une statistique permettant d'évaluer un paramètre inconnu (comme son espérance ou sa variance) relatif à une loi de probabilité. un estimateur est une fonction qui fait correspondre à chaque réalisation possible x_1, \dots, x_n de l'échantillon à n éléments $\hat{\theta} = f(x_1, x_2, \dots, x_n)$
- ❖ Un « Biais » est procédé qui engendre des erreurs dans les résultats d'une étude. C'est une erreur utilisée pour définir la qualité d'un algorithme d'apprentissage. En statistique, c'est une variable aléatoire fluctue autour de son espérance. On peut donc souhaiter que l'espérance de $\hat{\theta}$ soit égale à θ c.-à-d. en « **moyenne** » l'estimateur ne se trompe pas : $\text{Biais}(\hat{\theta}) = \mathbb{E}(\hat{\theta}) - \theta$

=> Si le biais est égal à zéro, l'**estimateur** est dit sans biais.



Éléments de statistiques

I. Qu'est-ce que Estimateur non biaisé ?

- ❖ La qualité des estimateurs s'exprime par leur **convergence**, leur **biais**, leur **efficacité** et leur **robustesse**.
- ❖ Une **variable aléatoire certaine** X_C est une variable qui ne prend **qu'une seule valeur**, avec $\mathbb{P}(X_C) = 1$. On a $V(X_C) = 0 \Leftrightarrow X_C$ est une **v.a.r. certaine**, $\mathbb{E}(X_C) = X_C$
 \Rightarrow . Une **variable aléatoire certaine** X_C **a un estimateur non biaisé**

Pour faire **l'apprentissage d'un modèle**, il faut chercher **des estimateurs non biaisés** des paramètres d'un modèle, \Rightarrow Il faut que ces derniers soient considérés comme des **variables certaines**. C'est cette approche, dite **fréquentiste**

- ✓ De même pour l'approche **bayésienne** qui considère les paramètres du modèle comme des variables aléatoires, permet également d'obtenir **des excellents résultats**



Éléments de statistiques

➤ L'approche probabiliste «fréquentiste»

- ❖ Tous les algorithmes de machine learning sont basés sur l'approche fréquentiste
- ❖ Considérons un jeu de données et un ensemble possible de distributions au sens probabiliste, à savoir les valeurs que peuvent prendre les variables aléatoires et à quelle fréquence. Les approches probabilistes permettent de classer les variables :

=> **L'approche fréquentiste**, repose sur les observations et consiste à trouver la distribution la plus probable au vue des données, et éventuellement celle que son intervalle de confiance correspondant **(c'est à dire l'ensemble des distributions qui ont une chance significative d'être la vraie distribution)**. Cette approche est basée sur la « loi des grands nombres » et la théorie des « inégalités de concentration ».

- ❖ Parmi les difficultés majeures **des approches fréquentistes** est le problème du **surapprentissage** « Overfitting » qui est généralement contourné en utilisant une régularisation, un outil à la fois riche, flexible mais complexe à utiliser.



Éléments de statistiques

➤ L'approche probabiliste «fréquentiste»

✓ « Overfitting » :

Si un **algorithme d'apprentissage supervisé**, cherche le modèle qui exprime le mieux la relation entre des données. Et il apprend à partir des données mais aussi à partir de **patterns (schémas, structures)** qui ne sont pas liés au problème, comme du bruit. On dit que l'algorithme **surapprend**. L'overfitting est caractérisé par une erreur de type variance très élevée. Concrètement, on observe généralement de l'overfitting lorsqu'on utilise des modèles très complexes sur des problèmes simples mais bruités

✓ « Régularisation » :

C'est la technique clé de machine learning qui vise à limiter le (overfitting) et à contrôler l'erreur de type variance pour aboutir à des meilleures performances. La régularisation permet d'imposer une contrainte pour favoriser les modèles simples au détriment des modèles complexes. Autrement dit, cela permet de réduire l'erreur de type variance et d'améliorer la généralisation de la solution.



Éléments de statistiques

➤ L'approche probabiliste «fréquentiste»

✓ « Loi des grands nombres » :

La loi des grands nombres permet d'interpréter la probabilité comme une fréquence de réalisation, justifiant ainsi le principe des sondages, et présente l'espérance comme une moyenne. Plus formellement, elle signifie que la moyenne empirique, $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$ calculée sur les valeurs d'un échantillon, converge vers l'espérance lorsque la taille de l'échantillon tend vers l'infini. $\mathbb{P}\left(\lim_{n \rightarrow +\infty} \bar{X}_n = E(X)\right) = 1$

✓ « Inégalité de concentration » :

Les **inégalités de concentration** fournissent des bornes sur la probabilité qu'une statistique basée sur des tirages successifs d'une variable aléatoire dévie d'une certaine valeur. Parmi les plus connus, il y'a **Azuma- Hoeffding**. Elles permettent de contrôler l'incertitude liée aux quantités statistiques comme la moyenne empirique. Par exemple en utilisant l'inégalité **d'Azuma-Hoeffding**, on sait que si on lance 10 000 fois une pièce équilibrée, la proportion de tirages côté pile sera comprise entre 0,483 et 0,517 avec une probabilité supérieure à 99 %. Ces résultats permettent de calculer des intervalles de confiance autour des prédictions des modèles de machine learning.



Éléments de statistiques

➤ L'approche probabiliste «fréquentiste»

❖ Exemple:

Une pièce de monnaie a une probabilité qu'elle tombe sur pile ou face. **L'approche fréquentiste** se basera sur l'expérience présente (par exemple, la pièce, lancée **10** fois, et tombée 6 fois sur pile) pour établir que la « vraie » probabilité d'obtenir pile est $6/10=0,6$. Si on lance la pièce en un nombre important de fois, selon « **loi des grands nombres** », cette méthode convergera, au sens mathématique, vers **0,5**. Concrètement, on utilise des « **inégalités de concentration** » pour quantifier l'incertitude du résultat autour de bornes de probabilités. Et si on lance une pièce équilibrée **10 000** fois, la proportion de pile sera comprise entre **0,483 et 0,517** avec une probabilité supérieure à **99 %**.



Éléments de statistiques

➤ L'approche probabiliste «bayésienne»

- ❖ **L'approche probabiliste bayésienne** c'est une approche, que l'on qualifie parfois de théorique ou déductive, combine l'information apportée par les données avec les connaissances a priori provenant soit d'études antérieures soit d'avis d'experts, dans le but d'obtenir une information a posteriori.

✓ Exemple:

Prenons l'exemple précédent : lançons une pièce 10 fois. Si on se base sur une modélisation fréquentiste, alors il existe une « vraie » probabilité d'obtenir pile, qui a comme valeur p . Si on obtient 6 fois sur 10 pile, alors la probabilité d'obtenir pile à partir des résultats de cette expérience est égale à $6/10 = 0.6$. Selon l'approche bayésienne, on ne s'intéresse pas à cette probabilité mais plutôt à sa loi a priori. En effet, si la pièce est équilibrée alors a priori la probabilité d'obtenir pile est la même que celle d'obtenir face, c'est-à-dire $1/2 = 0.5$. Cette probabilité a priori est obtenue à partir des résultats d'autres expériences effectuées par le passé.



Éléments de statistiques

➤ L'approche probabiliste «bayésienne»

❖ L'utilité de l'approche bayésienne

L'intérêt de cette approche est donc fort quand on peut tenir compte d'expériences passées parfaitement similaires. Elle est donc utilisée dans plusieurs domaines comme par exemple la détection de spams : la connaissance préalable des spams permet d'associer une probabilité correspondant au nombre de fois où un type de mot apparaît. Cette probabilité, obtenue grâce aux expériences passées, permet de considérer un mot comme étant typique d'un spam. Le principal avantage de la méthode est donc de s'affranchir d'un horizon fixe et d'avoir des résultats le plus rapidement possible. Inutile de fixer au préalable la taille d'un échantillon nécessaire et d'un niveau de trafic pour effectuer un test : les résultats sont consultables tout au long de l'expérience et sont plus rapides à obtenir.



Éléments de statistiques

I. Qu'est-ce que Estimateur non biaisé ?

- ❖ **La moyenne est un estimateur non biaisé de l'espérance mathématique**
- ❖ Il reste à évaluer la qualité de cette estimation : le fait qu'elle soit non biaisée ne garantit pas qu'elle soit précise : sa précision dépend du nombre et de la « qualité » des mesures effectuées, c'est-à-dire de la dispersion des mesures autour de l'espérance mathématique.

=> Pour caractériser numériquement cette dispersion, on utilise la notion de **Variance**.

J. Qu'est-ce que la Variance ?

C'est une erreur utilisée pour définir la qualité d'un algorithme d'apprentissage, cet algorithme tente d'approcher la relation exacte entre des variables d'entrée et de sortie d'un problème, comme étant **le « vrai modèle »**. Il peut être tentée pour 'utiliser le modèle le plus complet et complexe afin d'être sûr de capturer toutes les subtilités d'un problème. Malheureusement les données sont souvent bruitées, c'est à dire altérées par la présence d'un signal aléatoire de faible intensité



Éléments de statistiques

J. Qu'est-ce que la Variance ?

Ce bruit ne contient aucune information et il ne peut pas être séparé des données. Par exemple les bruits d'image qui altèrent la valeur des pixels. Lors, utilisation d' un modèle trop compliqué revient à essayer d'apprendre la structure du bruit – une structure qui n'existe pas -, ce qui donne lieu à d'améliorer les performances par la « Régularisation » afin d'équilibrer les erreurs de la variance.

On dit aussi que la variance représente la « distance » entre le vrai modèle et la solution à laquelle l'algorithme est arrivé. Cette erreur est fortement influencée par la complexité du modèle, celle du bruit, et la quantité de données disponibles. La variance peut être très élevée si le modèle est beaucoup trop compliqué par rapport au bruit et au nombre de données

Définition

La variance d'une variable aléatoire Y de distribution $f_Y(y)$ est la quantité

$$Var(Y) = \sigma^2 = \int_{-\infty}^{+\infty} [y - \mathbb{E}(Y)]^2 f_Y(y) dy$$



Éléments de statistiques

J. Qu'est-ce que la Variance ?

=> L'écart type σ est une mesure de la dispersion des valeurs d'un échantillon statistique ou d'une distribution de probabilité $\sigma = [\text{Var}(Y)]^{1/2}$

Propriété

- ✓ Une variable certaine a une variance nulle
- ✓ $\text{Var}(Y) = \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2$
- ✓ $\text{Var}(aY) = a^2 \text{Var}(Y)$
- ✓ Si une variable aléatoire obéit à une distribution uniforme sur un intervalle $[a, b]$, sa variance vaut $(b-a)^2/12$.
- ✓ Si une variable aléatoire obéit à une loi gaussienne d'écart-type σ , sa variance vaut σ^2

J. Qu'est-ce que la Covariance des deux variables aléatoires ?

Définition

La covariance de deux variables aléatoires U et V est définie par :

$$\text{Cov}_{U,V} = \mathbb{E}[(U - \mathbb{E}(U))(V - \mathbb{E}(V))] = \mathbb{E}[UV] - \mathbb{E}[U]\mathbb{E}[V]$$

Éléments de statistiques

□ Exemples des lois de probabilité

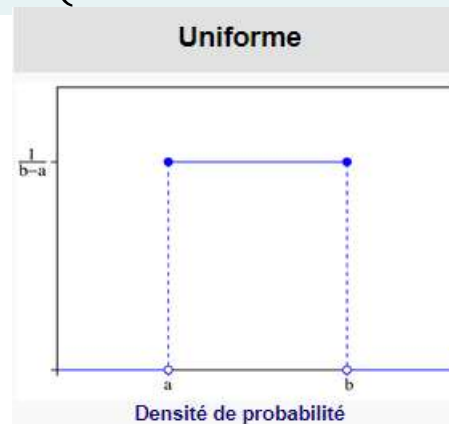
i. La loi uniforme

La loi uniforme a la propriété suivante tous les intervalles de même longueur inclus dans la loi ont la même probabilité. Cela se traduit par le fait que la densité de probabilités de ces lois est constante.

Définition

La densité de probabilité de la loi uniforme continue est une fonction indicatrice sur l'intervalle $[a, b]$:

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





Éléments de statistiques

□ Exemples des lois de probabilité

ii. La loi gaussienne

La loi gaussienne est parmi les lois de probabilité les plus adaptées pour modéliser des phénomènes naturels issus de plusieurs événements aléatoires. C'est une loi de probabilité paramétrique caractérisée par sa moyenne μ et sa variance σ^2 , $X \sim \mathcal{N}(\mu, \sigma^2)$.

si est seulement si il a cette densité de probabilité :

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2}.$$

Cette fonction de densité, appelée courbe de Gauss, possède une forme caractéristique rappelant celle d'une cloche. Elle se répartit de manière symétrique autour de la moyenne, point où elle atteint son maximum. Elle décroît ensuite à mesure que les valeurs sont éloignées de la moyenne. Cette concentration de la densité autour de la moyenne est une caractéristique importante de la loi gaussienne. Elle se traduit notamment par le fait qu'une majorité des valeurs observées est contenue dans **un intervalle restreint autour de la moyenne**.

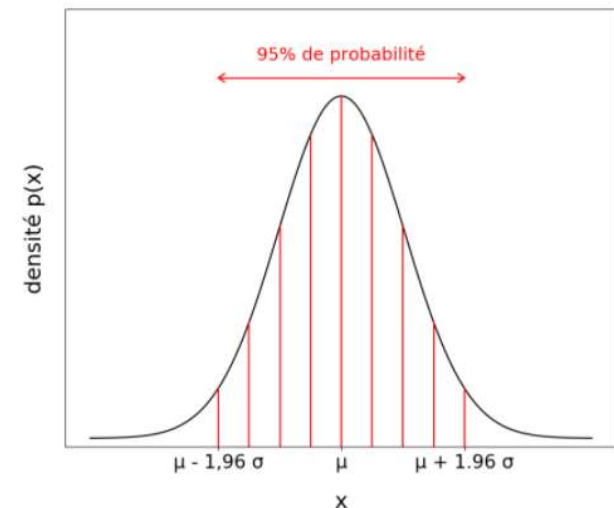
Éléments de statistiques

□ Exemples des lois de probabilité

ii. La loi gaussienne

Par exemple, la probabilité d'avoir une valeur comprise dans l'intervalle $[\mu - 1.96 \sigma, \mu + 1.96 \sigma]$ est égale à **95 %**

la loi gaussienne est à la base de nombreuses méthodes d'apprentissage statistique, comme les tests statistiques et les modèles utilisant les processus gaussien.



La loi gaussienne de moyenne nulle et d'écart-type égal à 1 est appelée **loi gaussienne centrée réduite** ou **loi gaussienne standard**. elle correspond au comportement, sous certaines conditions, d'une suite d'expériences aléatoires similaires et indépendantes. lorsque le nombre d'expériences est très élevé. Grâce à cette propriété, une loi gaussienne permet d'approcher à d'autres lois et ainsi de modéliser de nombreuses études scientifiques comme des mesures d'erreurs ou des tests statistiques,