Rappels sur le calcul des probabilités et la statistique

Introduction

- La théorie des probabilités est l'outil idéal pour formaliser les hypothèses et les incertitudes par rapport au données.

- Les données seront traitées comme des variables aléatoires.

✓ **La valeur d'une variable aléatoire est incertaine avant de l'observer.**

✓ **La loi de probabilité d'une variable aléatoire caractérise l'incertitude par rapport à sa valeur.**

Expérience aléatoire et événements

- Un espace d'échantillons (universel) Ω est l'ensemble des résultats possibles d'une expérience aléatoire.

  Ex. Lancer une pièce de monnaie 2 fois, $\Omega = \{\mathrm{PP}, \mathrm{PF}, \mathrm{FP}, \mathrm{FF}\}$

- Un évènement est un sous-ensemble de Ω :

  Ex. La première lancée est une pile: $\Omega = \{\mathrm{PP}, \mathrm{PF}\}$.

- S : espace d'événements :
  - ✓ Est fermé sous les opérations d'union et de complément.
  - ✓ Admet les opérations d'ensembles: union, différence, etc.
  - ✓ Contient l'ensemble vide $\emptyset$ et l'événement certain Ω.

Probabilités sur les événements

Définir sur $(\boldsymbol{\Omega}, \boldsymbol{S})$:

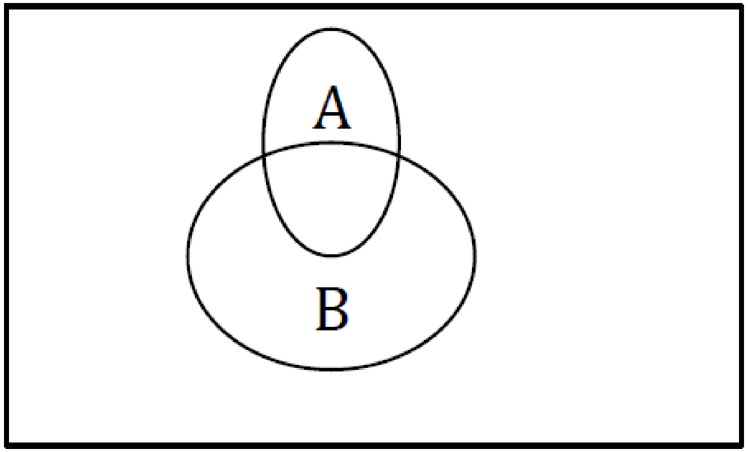
- $p(A) \geq 0$ pour chaque A dans S.
- $p(\Omega) = 1.$
- $P(A \ U \ B) = p(A) + p(B) - p(A \cap B)$
- Si A, B sont 2 événements disjoints: $p(A \cap B) = 0$, alors:

$$p(A \ U \ B) = p(A) + p(B)$$

- $p(\neg A) = 1 - p(A)$.

Probabilité conditionnelle

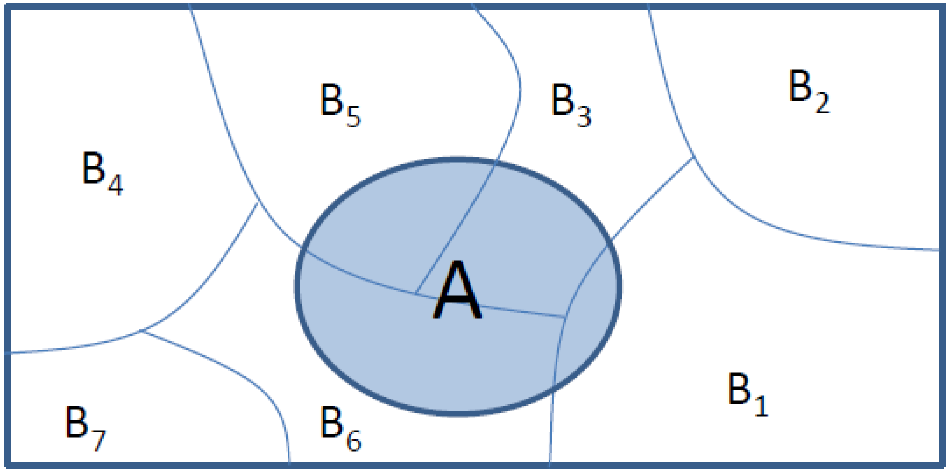
- $p(A|B)$ = Fraction où A est vrai sachant que B est vrai aussi. On a alors la **loi de Bayes**:



$$p(\mathrm{A} \mid \mathrm{B}) = \frac{p(\mathrm{A} \cap \mathrm{B})}{p(\mathrm{B})} \quad \Rightarrow \quad p(\mathrm{A} \cap \mathrm{B}) = p(\mathrm{A} \mid \mathrm{B}) p(\mathrm{B})$$

Probabilité marginale

- Soit A un événement et $\{B_1, B_2, \ldots, B_n\}$ une partition de Ω.



- On a la probabilité marginale:

$$p(A) = \sum_{i=1}^{n} p(A \cap B_i) = \sum_{i=1}^{n} p(B_i) p(A|B_i)$$

Les variables aléatoires

Manière concise de spécifier les résultats d'expérience.

- **Ce sont des valeurs numériques continues ou discrètes.**

- VA discrètes prennent des valeurs dénombrables distinctes.

  Exemple: - *Nombre de dinars dans mon compte bancaire.*

  - Nombre total de face après 100 lancée d'une pièce.
  (valeurs possibles: 0, 1, 2, …, 100).

- VA continues prennent des valeurs non-dénombrables (réelles):

  Exemple: *Quantités physiques: résistance de l'air, forces, etc.*

Règles de probabilités

Soit x et y deux variables aléatoires discrètes:

x peut prendre les valeurs: $u_1, u_2, \ldots, u_N.$

y peut prendre les valeurs: $v_1, v_2, \ldots, v_M$

On aura alors:

- $\sum_{i=1}^{N} p(\mathrm{x} = u_i)=1$
- $\sum_{j=1}^{M} p(\mathrm{y} = v_j)=1$
- La probabilité conjointe d'observer $\mathrm{x} = u_i$ et $\mathrm{y} = v_j$ est :

$$\boxed{p(\mathrm{x} = u_i , \mathrm{y} = v_j)}$$

Règles de probabilités

Probabilité marginale

Lorsqu'on s'intéresse à la probabilité d'une variable aléatoire et pas au autres variables existantes, on défini sa probabilité marginale, comme suit:

Ex. Soit x et y deux variables aléatoires discrètes, qui prennent les valeurs $\{u_1, u_2, \ldots, u_N\}$ et $\{v_1, v_2, \ldots, v_M\}$, la probabilité marginale de x est donnée par:

$$p(\mathrm{x} = u_i) = \sum_{j=1}^{M} p(\mathrm{x} = u_i, \mathrm{y} = v_j)$$

Règles de probabilités

Probabilité conditionnelle

- Lorsqu'on s'intéresse à la probabilité d'une variable aléatoire sachant qu'on a observé une ou d'autres variables existantes, on défini sa probabilité conditionnelle, comme suit:

Ex. La probabilité d'observer $\mathrm{y} = v_j$ sachant $\mathrm{x} = u_i$ que est:

$$p(\mathrm{y} = v_j | \mathrm{x} = u_i) = \frac{p(\mathrm{x} = u_i | \mathrm{y} = v_j) p(\mathrm{y} = v_j)}{p(\mathrm{x} = u_i)}$$

- Utile lorsqu'on veut raisonner sur y, après avoir observé x.

Règles de probabilités

Règle du produit

- La probabilité conjointe de deux variables aléatoire peut toujours être décomposée en une probabilité marginale et une probabilité conditionnelle.

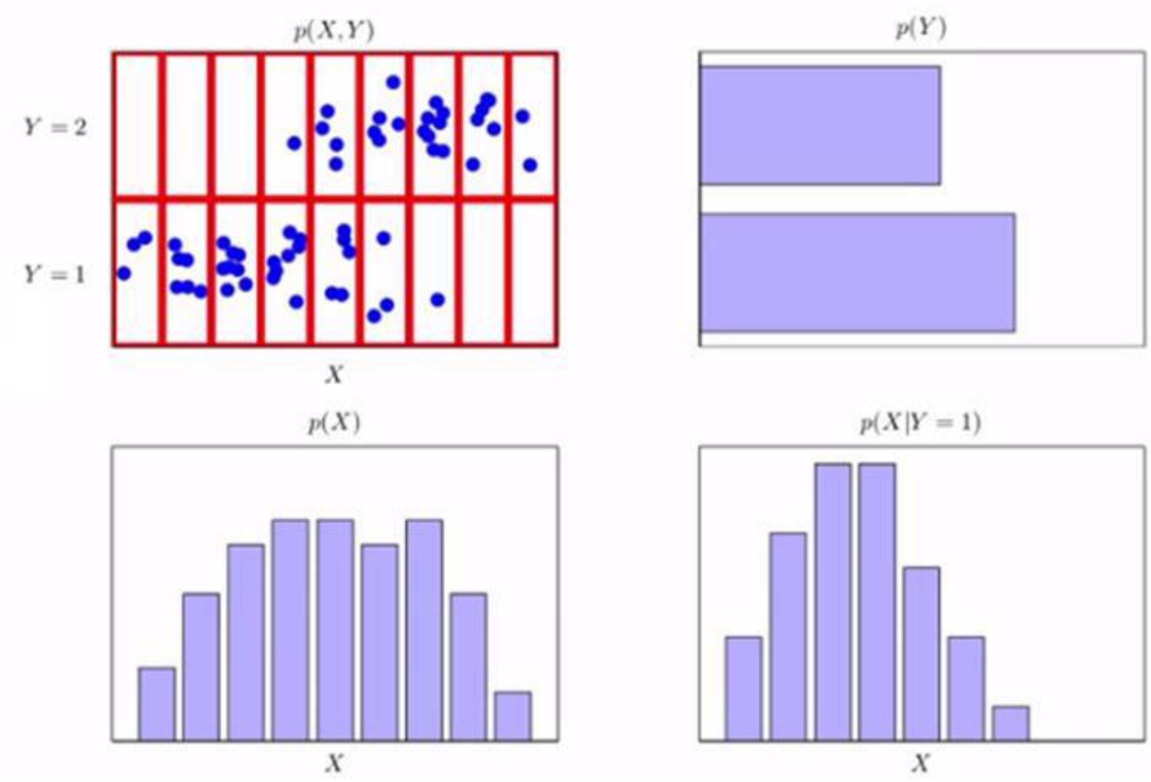
$$p(\mathrm{x} = u_i, \mathrm{y} = v_j) = p(\mathrm{x} = u_i | \mathrm{y} = v_j) p(\mathrm{y} = v_j)$$

- **En mots:**

La probabilité d'observer $\mathrm{x} = u_i$ et $\mathrm{y} = v_j$ en même temps, c'est la probabilité d'observer $\mathrm{y} = v_j$ multiplié par la probabilité d'observer $\mathrm{x} = u_i$ sachant que $\mathrm{y} = v_j$ est observé.

Règles de probabilités

Probabilité jointe, marginale et conditionnelle



Règles de probabilités

Règle de Bayes pour les variables aléatoires

- Permet d'inverser l'ordre des probabilités conditionnelles.

$$p(\mathrm{y} = v_j | \mathrm{x} = u_i) = \frac{p(\mathrm{x} = u_i | \mathrm{y} = v_j) p(\mathrm{y} = v_j)}{p(\mathrm{x} = u_i)}$$

☞ $p(\mathrm{y} = v_j)$ est appelée probabilité a priori.

☞ $p(\mathrm{y} = v_j | \mathrm{x} = u_i)$ est appelée probabilité a posteriori.

Règles de probabilités

Indépendance de variables aléatoires

- Deux variables aléatoires x et y sont indépendantes si:

☞ $p(\mathrm{x} = u_i, \mathrm{y} = v_j) = p(\mathrm{x} = u_i)p(\mathrm{y} = v_j).$

Ou

☞ $p(\mathrm{y} = v_j | \mathrm{x} = u_i) = p(\mathrm{y} = v_j)$

Ou

☞ $p(\mathrm{x} = u_i | \mathrm{y} = v_j) = p(\mathrm{x} = u_i).$

Distributions de variables aléatoires

Loi de Bernoulli $x \sim Ber(\theta)$

- x est une variable binaire $x \in \{0, 1\}$.
- $p(x = 1) = \theta$, $p(x = 0) = 1 - \theta$.

Loi binomiale $x \sim Bin(n, \theta)$

- La somme de variables de Bernoulli;
- x peut prendre les valeurs: 0, 1, ..., n.

Ex. Lancer une pièce de monnaie n fois et compter le nombre de piles (succès) obtenus x.

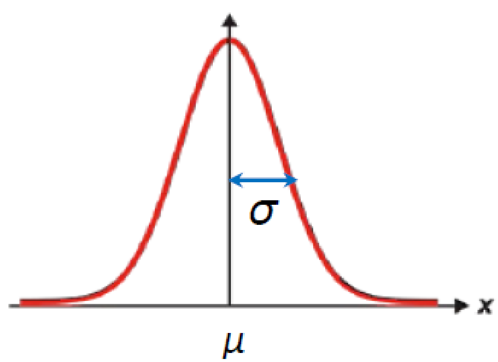
$$p(x = u) = \binom{n}{u} \theta^u (1 - \theta)^{n-u} \qquad \binom{n}{u} = \frac{n!}{(n-u)!\,u!}$$

Distributions de variables aléatoires

Loi normale (Gaussienne) à 1 dimension $\mathrm{x} \sim \mathcal{N}(\mu, \sigma^2)$

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

Ex: La taille d'une population x d'une ville quelconque suit une loi normale $\mathcal{N}(\mu{=}168\ cm,\ \sigma^2{=}100)$,



Moyenne d'une variable aléatoire (VA)

- On dénote la moyenne (espérance) d'une VA x par: $\mu = \mathbb{E}(\mathrm{x})$.

  ☞ Pour VA discrète: $\mathbb{E}(\mathrm{x}) = \sum_{i=1}^{N} u_i \, p(\mathrm{x} = u_i)$

  ☞ Pour V-A continue: $\mathbb{E}(\mathrm{x}) = \int_{-\infty}^{+\infty} u \, f(\mathrm{x} = u) \, du$

- Si g est une fonction de x , alors:

  ☞ Pour VA discrète: $\mathbb{E}(g(\mathrm{x})) = \sum_{i=1}^{N} g(u_i) \, p(\mathrm{x} = u_i)$

  ☞ Pour V-A continue: $\mathbb{E}(g(\mathrm{x})) = \int_{-\infty}^{+\infty} g(u) \, f(\mathrm{x} = u) \, du$

Moyenne d'une variable aléatoire (VA)

Propriétés:

Soit x et y deux variables aléatoires et α une constante réelle.

- ☞ $\mathbb{E}\,(\mathrm{x}+\mathrm{y}) = \mathbb{E}\,(\mathrm{x}) + \mathbb{E}\,(\mathrm{y})$.

- ☞ $\mathbb{E}\,(\alpha\,\mathrm{x}) = \alpha\mathbb{E}\,(\mathrm{x})$.

- ☞ Si x et y sont indépendantes, alors: $\mathbb{E}\,(\mathrm{xy}) = \mathbb{E}\,(\mathrm{x})\mathbb{E}\,(\mathrm{y})$.

Moyenne d'une variable aléatoire (VA)

- On dénote la variance d'une variable x par: $\sigma^2 = \mathbb{E}[(\mathrm{x}-\mu)^2]$.

☞ Pour VA discrète: $\sigma^2 = \sum_{i=1}^{N} (u_i - \mu)^2 p(\mathrm{x} = u_i)$

☞ Pour VA continue: $\sigma^2 = \int_{-\infty}^{+\infty} (u-\mu)^2 f(\mathrm{x} = u) du$

Propriétés: Si $var(\mathrm{x}) = \sigma^2$ est la variance de x, alors:

☞ Pour α constante réelle: $var(\alpha\,\mathrm{x}) = \alpha^2 var(\mathrm{x})$.

☞ Si x et y sont deux VA indépendantes, alors:

$$var(\mathrm{x}+\mathrm{y}) = var(\mathrm{x}) + var(\mathrm{y}).$$

Covariance entre deux variables aléatoires

- Soit deux VA x et y discrètes dont les moyennes sont μ_{x} et μ_{y} et les variances ${\sigma_{\mathrm{x}}}^2$ et ${\sigma_{\mathrm{y}}}^2$, respectivement. La covariance entre les deux variables x et y est définie par:

$$cov(\mathrm{x},\mathrm{y}) = \sum_{i=1}^{N} \sum_{j=1}^{M} u_i\, v_j p(\mathrm{x} = u_i, \mathrm{y} = v_j) - \mu_{\mathrm{x}}\mu_{\mathrm{y}}$$

- Le coefficient de corrélation entre x et y est défini par

$$\rho(\mathrm{x},\mathrm{y}) = cov(\mathrm{x},\mathrm{y})/(\sigma_{\mathrm{x}}\sigma_{\mathrm{y}})$$

Propriétés:

☞ $-1 \leq \rho(\mathrm{x},\mathrm{y}) \leq 1.$ ☞ $cov(\mathrm{x},\mathrm{x}) = var(\mathrm{x}).$

Distributions Gaussienne multivariée

- Soit D variables aléatoires $\mathrm{x}_1, \mathrm{x}_2, \dots, \mathrm{x}_D$ et soit $\mu_1, \mu_2, \dots, \mu_D$ leurs moyennes, respectivement.

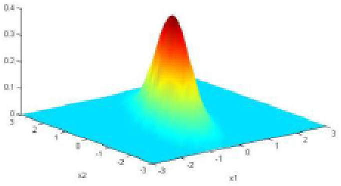
- On définit la matrice de covariance $\boldsymbol{\Sigma}$ de dimension $D \times D$ dont les entrées sont définies par: $\boldsymbol{\Sigma}_{i,j} = cov(\mathrm{x}_i, \mathrm{x}_j)$.

- On peut définir alors la loi Gaussienne multivariée: $\mathrm{x} \sim \mathcal{N}(\mu, \boldsymbol{\Sigma})$ où $\mathrm{x} = (\mathrm{x}_1, \mathrm{x}_2, \dots, \mathrm{x}_D)$ et $\mu = (\mu_1, \mu_2, \dots, \mu_D)$:

$$f(\mathrm{x} = u) = \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(u-\mu)^T \boldsymbol{\Sigma}^{-1}(u-\mu)\right)$$

Ex:

$$\mu = (0,0)^T \qquad \Sigma = \begin{pmatrix} 0.25 & 0.30 \\ 0.30 & 1.00 \end{pmatrix}$$



Références

1. M. S. Allili. Techniques d’apprentissage automatique (Cours de 2e cycle). Université du Québec en Outaouais (UQO), Québec, Canada. Hivers 2015.
2. S. Rogers et M Girolami. A first Course in machine learning, CRC press, 2012.
3. C. Bishop. Pattern Recognition and Machine learning. Springer 2006.
4. R. Duda, P. Storck et D. Hart. Pattern Classification. Prentice Hall, 2002.