Méthodes d'apprentissage

IFT603-712

Formulation probabiliste

Par

Pierre-Marc Jodoin

et

Hugo Larochelle

Illustration au tableau des probabilités **marginales**, **jointes** et **conditionnelles**

Variable aléatoire

- La théorie des probabilités est l’outil idéal pour formaliser nos **hypothèses et incertitudes** par rapport à nos données

- On va traiter nos données comme des **variables aléatoires**

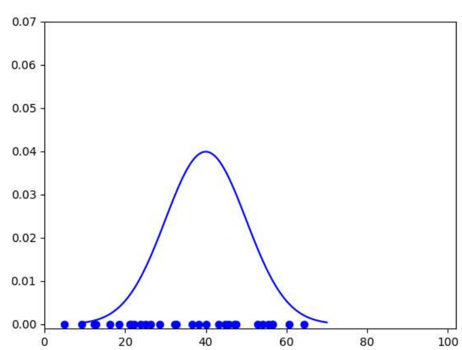
  - ➢ la valeur d’une variable aléatoire est incertaine (avant de l’observer)
  - ➢ la loi de probabilité de la variable aléatoire caractérise notre incertitude par rapport à sa valeur

Variable aléatoire

$$x \sim P(X)$$

Exemple, résultats à un examen (groupe 1)

$$x \sim N(\mu = 40, \sigma = 10)$$



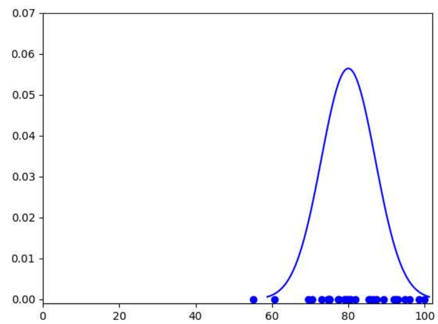
4

Variable aléatoire

$$x \sim P(X)$$

Exemple, résultats à un examen (groupe 2)

$$x \sim N(\mu = 80, \sigma = 7)$$



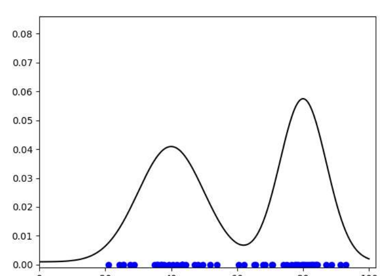
5

Variable aléatoire

$$x \sim P(X)$$

Exemple, résultats à un examen (toute la classe)

$$x \sim \frac{1}{2} N(\mu = 40, \sigma = 10) + \frac{1}{2} N(\mu = 80, \sigma = 7)$$



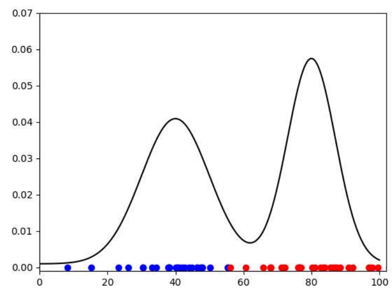
6

Variable aléatoire

$$x \sim P(X)$$

Exemple, résultats à un examen (toute la classe)

$$x \sim \frac{1}{2} N(\mu = 40, \sigma = 10) + \frac{1}{2} N(\mu = 80, \sigma = 7)$$



Données étiquetées

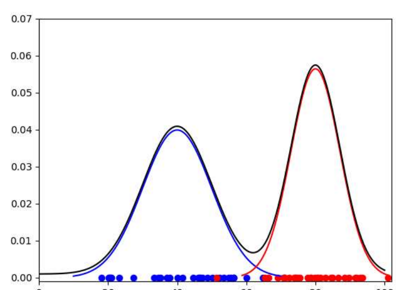
7

Variable aléatoire

$$x \sim P(X)$$

Exemple, résultats à un examen (toute la classe)

$$x \sim \frac{1}{2} N(\mu = 40, \sigma = 10) + \frac{1}{2} N(\mu = 80, \sigma = 7)$$



Résultat d'un entraînement

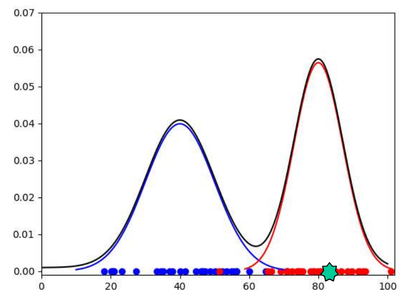
8

Variable aléatoire

$$x \sim P(X)$$

Exemple, résultats à un examen (toute la classe)

$$x \sim \frac{1}{2} N(\mu = 40, \sigma = 10) + \frac{1}{2} N(\mu = 80, \sigma = 7)$$



Très confiant que le nouvel étudiant ayant eu la note de 82% ( ) est issu du groupe 2

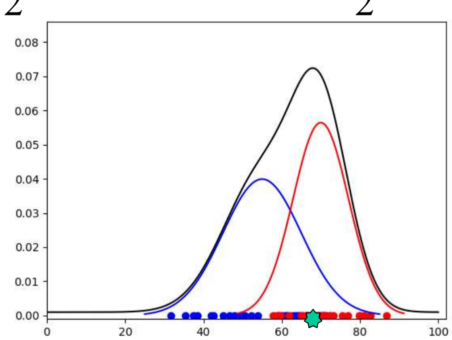
9

Variable aléatoire

$$x \sim P(X)$$

Exemple, résultats à un examen (toute la classe)

$$x \sim \frac{1}{2} N(\mu = 50, \sigma = 10) + \frac{1}{2} N(\mu = 70, \sigma = 6)$$



Incertain que le nouvel étudiant ayant eu la note de 67% ( ) est issu du groupe 2

10

Variable aléatoire

- Soient X et T des variables aléatoires **discrètes**
  - X peut prendre comme valeurs $x_1, \ldots, x_M$
  - T peut prendre comme valeurs $t_1, \ldots, t_M$

- La **probabilité jointe** qu'on observe $X = x_i$ et $T = t_j$ est notée

$$P(X = x_i, T = t_i)$$

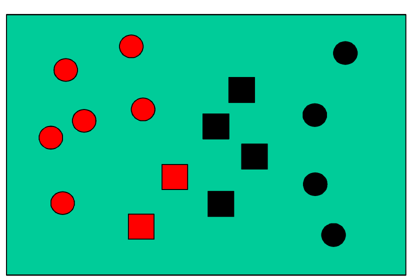
et se lit comme la « probabilité d'observer à **la fois** x_i **et** t_j .

- Note:

$$P(X = x_i, T = t_i) = P(T = t_i, X = x_i)$$

Probabilité jointe

Exemple $X : forme, \quad Y : couleur$



$$P(X = carré) = 6/16$$

$$P(Y = rouge) = 8/16$$

$$P(X = carré, Y = rouge) = 2/16$$

Probabilité marginale

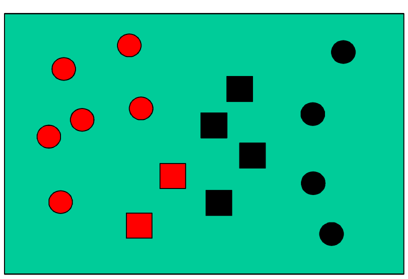
Une **probabilité marginale** est lorsqu'on ne s'intéresse pas à toutes les variables aléatoires qu'on a défini

Exemple : la probabilité marginale d'observer $X = x_i$

$$P(X = x_i) = \sum_{j=1}^{N} P(X = x_i, T = t_j)$$

Probabilité jointe

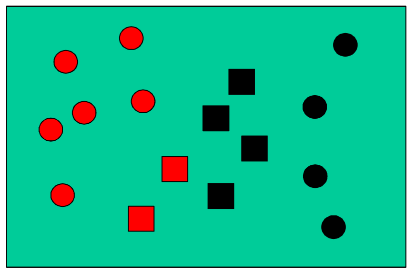
Exemple $X : forme, \quad Y : couleur$



$$P(X = carré) = \sum_{coleur} P(X = carré, Y = couleur)$$

Probabilité jointe

Exemple $X : forme, \quad Y : couleur$



$$P(X = carré) = P(X = carré, Y = rouge) + P(X = carré, Y = noir)$$
$$= \frac{2}{16} + \frac{4}{16}$$
$$= \frac{6}{16}$$

Probabilité conditionnelle

Une **probabilité conditionnelle** est lorsqu'on s'intéresse la valeur d'une variable aléatoire «étant donnée» une valeur assignée à d'autres variables

$$P(X = x_i \mid T = t_j)$$

Se lit : la probabilité que X= x_j étant donné que T= t_i

Probabilité conditionnelle

Exemple, élections américaines 2016

$$P(Voter\ républicain) = 46.1\%$$

VS

$$\begin{cases} P(Voter\ républicain \mid Zone\ urbaine) = 35\% \\ P(Voter\ républicain \mid Zone\ rurale) = 62\% \\ P(Voter\ républicain \mid Banlieu) = 50\% \end{cases}$$

https://www.npr.org/2016/11/14/501737150/rural-voters-played-a-big-part-in-helping-trump-defeat-clinton

Produit des probabilités

x_i et t_j ont disparu, seulement pour simplifier la notation

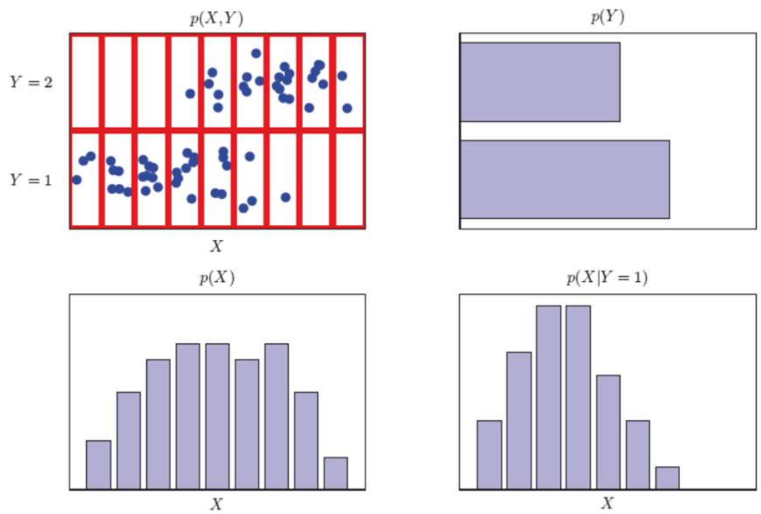
Une probabilité jointe peut toujours être décomposée par le produit d'une probabilité conditionnelle et marginale

$$P(X,T) = P(X \mid T)P(T)$$

En mots :

la probabilité d'observer $X = x_i$ **ET** $T = t_j$, c'est la probabilité d'observer $T = t_i$ multipliée par la probabilité d'observer $X = x_i$ **étant donné que** $T = xt_i$

Probabilités jointes, marginale et conditionnelles



Crédit : Bishop

Bayes

La **règle de Bayes** permet d'inverser l'ordre de la conditionnelle

$$P(T \mid X) = \frac{P(X \mid T)P(T)}{P(X)}$$

$p(T)$ est appelée loi de probabilité ***a priori*** (*prior*)

$p(T \mid X)$ est appelée loi de probabilité ***a posteriori*** (*posterior*)

Indépendance

Deux variables aléatoires X et T sont indépendantes si

- $P(X,T)=P(X)P(T)$ ou
- $P(X \mid T)=P(X)$ ou
- $P(T \mid X)=P(T)$

- Observer la valeur d'une variable ne nous apprend rien sur la valeur de l'autre

Variable aléatoire continue

Soit X une **variable aléatoire continue**

- X peut prendre un nombre infini de valeurs possibles (e.g. $\mathbb{R}$)
- X est associée à une **fonction de densité** de probabilité $p(x)$

la probabilité que X appartienne à un intervalle (a,b) est

$$p(x \in (a,b)) = \int_a^b p(x)dx$$

Variables aléatoires continues

Soit X une **variable aléatoire continue**

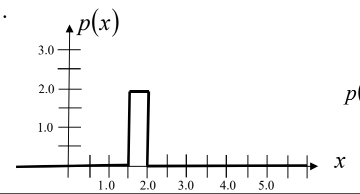
- la fonction de densité doit satisfaire

$$p(x) \geq 0$$

$$\int_{-\infty}^{\infty} p(x)dx = 1$$

à noter que, contrairement aux probabilités d'une variable discrète, la fonction de densité peut être > 1.

Exemple



$$p(x) = \begin{cases} 2 & \text{is } x \in [1.5, 2.0] \\ 0 & \text{sinon} \end{cases}$$

Variables aléatoires continues

Soit X une **variable aléatoire continue**

la **fonction de répartition** $P(z)$ (*cumulative distribution function*) donne la probabilité que X appartienne à l'intervalle $(-\infty, z)$

$$P(x = z) = \int_{-\infty}^{z} p(x)dx$$

Variables aléatoires continues

Soient X et T deux **variables aléatoires continues**

- elles sont associées à une **fonction de densité jointe** $p(x,t)$ telle que :

$$p(x \in [a,b], t \in [c,d]) = \int_a^b \int_c^d p(x,t)dxdt$$

Variables aléatoires continues

Soient X et T deux **variables aléatoires continues**

- La **function de densité marginale** s'obtient en intégrant l'autre variable

$$p(x) = \int p(x,t)dt$$

- La **function de densité conditionnelle** s'obtient comme auparavant

$$p(t \mid x) = \frac{p(x,t)}{p(x)}$$

Expérance mathématique

L'**espérance** d'une **variable *X*** est la moyenne qu'on obtient si on répète un grand nombre de fois une expérience

$$\mathrm{E}[X] = \sum_x x p(x) \quad \text{(cas discret)}$$

$$\mathrm{E}[X] = \int x p(x) dx \quad \text{(cas continu)}$$

Expérance mathématique

L'**espérance** d'une **fonction *f(x)*** est la moyenne qu'on obtient si on génère un grand nombre de valeurs pour cette fonction

$$\mathrm{E}[f] = \sum_x f(x) p(x) \quad \text{(cas discret)}$$

$$\mathrm{E}[f] = \int f(x) p(x) dx \quad \text{(cas continu)}$$

Variance

- La **variance** d'une **variable *X*** est

$$\mathrm{var}[X] = \mathrm{E}\left[\left(X - \mathrm{E}[X]\right)^2\right]$$

- La **variance** d'une **fonction *f (x)*** est

$$\mathrm{var}[f] = \mathrm{E}\left[\left(f(x) - \mathrm{E}[f(x)]\right)^2\right]$$

La variance mesure à quel point les valeurs varient autour de l'espérance

Propriétés de l'espérance et de la variance

Transformation linéaire de l'espérance

$$\mathrm{E}_{xy}[ax + by] = \sum_x \sum_y (ax + by)\,p(x, y)$$ a,b sont réels

$$= a\mathrm{E}[x] + b\mathrm{E}[y]$$ Si x, y indépendants

Transformation linéaire de la variance

$$\mathrm{var}[ax + by] = a^2\ \mathrm{var}[x] + b^2\ \mathrm{var}[y]$$ Si x, y indépendants

Espérance et variance conditionnelles

L'espérance et la variance se généralisent au cas **conditionnel** :

$$\mathrm{E}[x \mid y] = \sum_x x p(x \mid y)$$
$$\mathrm{E}[f(x) \mid y] = \sum_x f(x) p(x \mid y)$$

$$\mathrm{var}[x \mid y] = \mathrm{E}\left[\left(x - \mathrm{E}[x \mid y]\right)^2\right]$$
$$\mathrm{var}[f(x) \mid y] = \mathrm{E}\left[\left(f(x) - \mathrm{E}[f(x) \mid y]\right)^2\right]$$

Covariance

La covariance entre 2 variables aléatoires X et Y

$$\begin{aligned}\mathrm{cov}[x, y] &= \mathrm{E}_{xy}\left[(x - \mathrm{E}_x[x])(y - \mathrm{E}_y[y])\right] \\ &= \mathrm{E}_{xy}[xy] - \mathrm{E}_x[x]\mathrm{E}_y[y]\end{aligned}$$

mesure à quel point on peut prédire X à partir de Y (linéairement), et vice-versa

si X etY sont indépendantes, alors la covariance est 0

Variables aléatoires multidimensionnels

Une variable aléatoire peut être un vecteur

L'espérance d'un vecteur est le vecteur des espérances

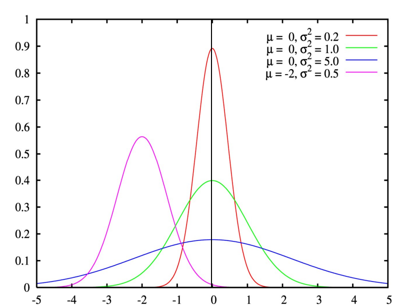
$$\mathrm{E}[\vec{x}] = \left(\mathrm{E}[x_1], \ldots, \mathrm{E}[x_D]\right)^{\mathrm{T}}$$

Et la covariance de deux vecteurs est

$$\mathrm{cov}[\vec{x}, \vec{y}] = \mathrm{E}_{\vec{x}\vec{y}}\left[\vec{x}\vec{y}^{\mathrm{T}}\right] - \mathrm{E}_{\vec{x}}[\vec{x}]\mathrm{E}_{\vec{y}}[\vec{y}]$$

Loi de probabilité gaussienne

$$N(x \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

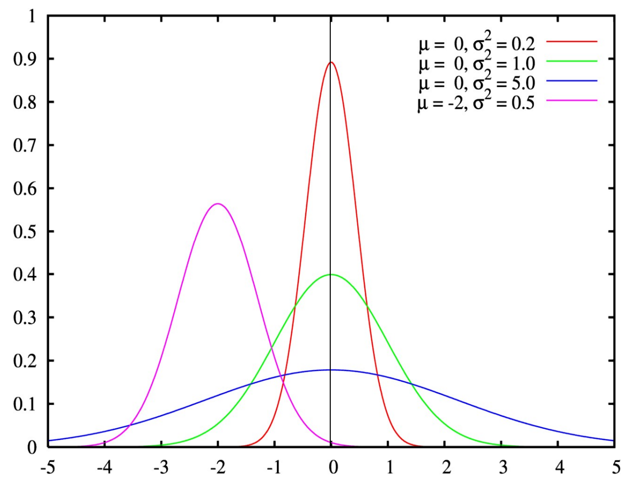


Moyenne : $\mathrm{E}[x] = \int_{-\infty}^{\infty} \mathrm{N}(x; \mu, \sigma) x dx = \mu$

Variance : $\mathrm{var}[x] = \int_{-\infty}^{\infty} \mathrm{N}(x; \mu, \sigma)(x-\mu)^2 dx = \sigma^2$

Écart type : $\sqrt{\mathrm{var}[x]} = \sigma$

Loi de probabilité gaussienne



σ	68.26894921371%
2σ	95.44997361036%
3σ	99.73002039367%

Gaussienne **multivariée**

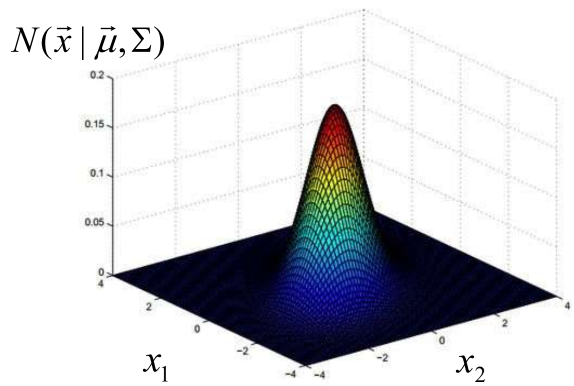
$$N(\vec{x} \mid \vec{\mu}, \Sigma) = \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} \exp\left\{ -\frac{1}{2} (\vec{x} - \vec{\mu})^{\mathrm{T}} \Sigma^{-1} (\vec{x} - \vec{\mu}) \right\}$$

Moyenne : $\mathrm{E}[\vec{x}] = \vec{\mu}$

Variance : $\mathrm{cov}[\vec{x}] = \Sigma$

Gaussienne **multivariée**

Exemple : $\vec{x} = (x_1, x_2)$

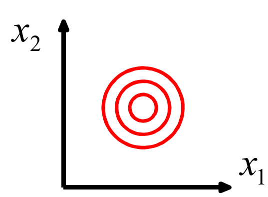


Gaussienne **multivariée**

Exemple : $\vec{x} = (x_1, x_2)$

Courbes de niveaux de $N(\vec{x} \mid \vec{\mu}, \Sigma)$

$$\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

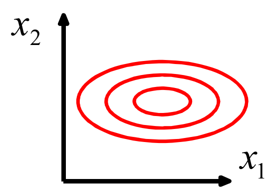


Gaussienne **multivariée**

Exemple : $\vec{x} = (x_1, x_2)$

Courbes de niveaux de $N(\vec{x} \mid \vec{\mu}, \Sigma)$

$$\Sigma = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$$

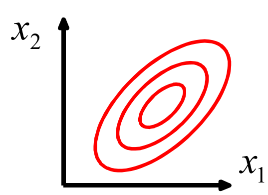


Gaussienne **multivariée**

Exemple : $\vec{x} = (x_1, x_2)$

Courbes de niveaux de $N(\vec{x} \mid \vec{\mu}, \Sigma)$

$$\Sigma = \begin{pmatrix} 1.5 & 0.5 \\ 0.5 & 1.5 \end{pmatrix}$$



Gaussienne **multivariée**

Une **combinaison linéaire** de variables aléatoires gaussiennes est également gaussienne

- Exemple
  - soit x une variable gaussienne de moyenne μ_1 et variance $\sigma_1{}^2$
  - soit t une variable gaussienne de moyenne μ_2 et variance $\sigma_2{}^2$
  - alors $ax + bt$ suit une loi gaussienne de moyenne $a\mu_1 + b\mu_2$ et variance $a^2\sigma_1{}^2 + b^2\sigma_2{}^2$ (x et y sont indépendantes)

Introduction au tableau

Maximum de vraisemblance

vs

Maximum a posteriori

43

Théorie de l'information

Théorie de l'information

- Les probabilités sont également utiles pour **quantifier l'information** présente dans des données

  exemple : quel est le nombre minimum de bits nécessaire pour encoder un message ?

- Cette question est intimement liée à la **probabilité d'observer ce message**

  plus le message est «surprenant» (improbable), plus on aura besoin de bits

Théorie de l’information

Codage de Huffman :

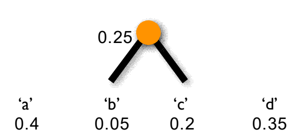
- façon optimale d’encoder des symboles indépendants de façon binaire
- plus un symbole est «fréquent» (probable), plus son code sera court

‘a’	‘b’	‘c’	‘d’
0.4	0.05	0.2	0.35

Théorie de l’information

Codage de Huffman :

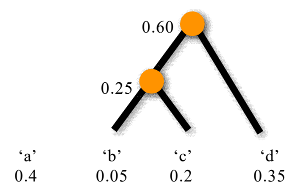
- façon optimale d’encoder des symboles indépendants de façon binaire
- plus un symbole est «fréquent» (probable), plus son code sera court



Théorie de l’information

Codage de Huffman :

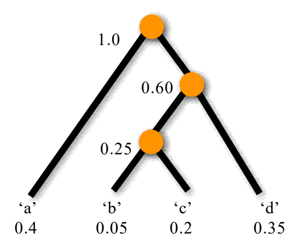
- façon optimale d’encoder des symboles indépendants de façon binaire
- plus un symbole est «fréquent» (probable), plus son code sera court



Théorie de l’information

Codage de Huffman :

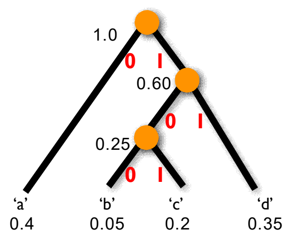
- façon optimale d’encoder des symboles indépendants de façon binaire
- plus un symbole est «fréquent» (probable), plus son code sera court



Théorie de l'information

Codage de Huffman :

- façon optimale d'encoder des symboles indépendants de façon binaire
- plus un symbole est «fréquent» (probable), plus son code sera court



Symbole	Code	Prob
'a'	0	40%
'b'	100	5%
'c'	101	20%
'd'	11	35%

Entropie

Symbole	Code	Prob
'a'	0	40%
'b'	100	5%
'c'	101	20%
'd'	11	35%

- Soit $p(x)$ la probabilité d'observer le symbole x

  la taille moyenne du code d'un symbole est

  $0.4 \times 1 + 0.05 \times 3 + 0.2 \times 3 + 0.35 \times 2 =$ **1.85 (bits)**

- **Entropie** :

$$H[x] = -\sum_x p(x) \log_2 p(x) \approx \underline{\mathbf{1.739\ (bits)}}$$

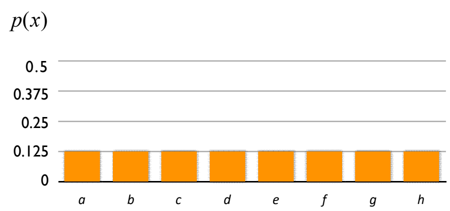
  Claude Shannon a démontré qu'il est impossible de compresser l'information dans un plus petit code moyen **sans perte d'information**

  $-\log_2 p(x)$ est l'information contenue par x

Entropie

Plus $p(x)$ est proche d'une **loi uniforme**, plus **l'entropie est élevée**

exemple : $x \in \{a,b,c,d,e,f,g,h\}$

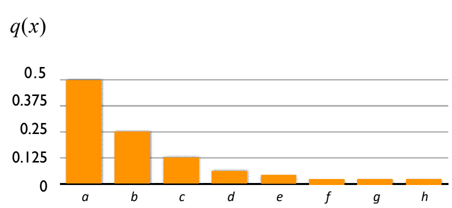


$$\begin{aligned}
H[x] &= -\sum_x p(x)\log_2 p(x) \\
&= -\left(\frac{1}{8}\log_2\left(\frac{1}{8}\right)\right)-\left(\frac{1}{8}\log_2\left(\frac{1}{8}\right)\right)-\left(\frac{1}{8}\log_2\left(\frac{1}{8}\right)\right)-\left(\frac{1}{8}\log_2\left(\frac{1}{8}\right)\right)-\left(\frac{1}{8}\log_2\left(\frac{1}{8}\right)\right)-\left(\frac{1}{8}\log_2\left(\frac{1}{8}\right)\right)-\left(\frac{1}{8}\log_2\left(\frac{1}{8}\right)\right)-\left(\frac{1}{8}\log_2\left(\frac{1}{8}\right)\right) \\
&= -8\left(\frac{1}{8}\log_2\left(\frac{1}{8}\right)\right) \\
&= 3 \text{ bits}
\end{aligned}$$

Entropie

Plus $q(x)$ est proche d'une **loi uniforme**, plus **l'entropie est élevée**

exemple : $x \in \{a,b,c,d,e,f,g,h\}$



$$\begin{aligned}
H[x] &= -\sum_x q(x)\log_2 q(x) \\
&= -\left(\frac{1}{2}\log_2\left(\frac{1}{2}\right)\right)-\left(\frac{1}{4}\log_2\left(\frac{1}{4}\right)\right)-\left(\frac{1}{8}\log_2\left(\frac{1}{8}\right)\right)-\left(\frac{1}{16}\log_2\left(\frac{1}{16}\right)\right)-\left(\frac{1}{32}\log_2\left(\frac{1}{32}\right)\right)-3\left(\frac{1}{64}\log_2\left(\frac{1}{64}\right)\right) \\
&= 2.06 \text{ bits}
\end{aligned}$$

Entropie

L'entropie se généralise aux variables continues

$$H[x] = -\int_{-\infty}^{\infty} p(x)\log_2 p(x)dx$$

Entropie relative et divergence de Kullback-Leibler

- Si on ne connaît pas $p(x)$, on va vouloir l'estimer
- Si $q(x)$ est notre estimation, on définit la **divergence de Kullback-Leibler** (K-L) comme suit :

$$KL\big(p(x)\,\|\,q(x)\big) = -\sum_x p(x)\log_2 q(x) - \left(-\sum_x p(x)\log_2 p(x)\right)$$

$$= \sum_x p(x)\log_2 \frac{p(x)}{q(x)}$$

> correspond au <u>**nombre de bits additionnels**</u> par rapport à ce qui serait optimal

Entropie jointe

L'entropie est une fonction d'une loi de probabilité

- elle reflète **l'incertitude représentée par la loi**
- si $p(x) = 1$ pour une seule valeur de x, l'entropie est 0

On peut généraliser l'entropie à **plusieurs variables**

$$H[x,y] = -\sum_x \sum_y p(x,y) \log_2 p(x,y)$$

Entropie conditionnelle

L'entropie conditionnelle quantifie **l'information additionnelle** qu'apporte une **nouvelle observation *y***

$$H[x \mid y] = -\sum_x \sum_y p(x,y) \log_2 p(x \mid y)$$

On peut démontrer que

$$H[x,y] = H[y \mid x] + H[x]$$

Information mutuelle

- Mesure à quel point deux variables sont indépendantes

$$I(x,y) = KL\big(p(x,y)\,||\,p(x)p(y)\big)$$
$$= \sum_{x,y} p\big(x,y\big)\log_2 \frac{p(x,y)}{p(x)p(y)}$$

- On appelle cette mesure l'**information mutuelle**

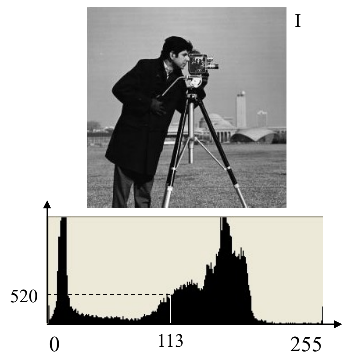
Autre exemple concret de l'utilité des probabilités

73

Comprendre la théorie des probabilité à l’aide d’images

Un **histogramme** représente le nombre de pixels appartenant à chaque niveau de gris (ou couleur) pouvant être représenté dans l’image.

$$H(c) = \text{Nb pixels d'intensité "}c\text{"}$$



74

Quelques définitions

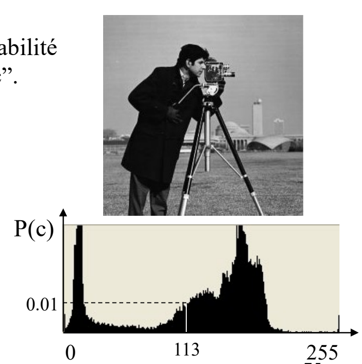
Parfois l’histogramme est normalisé par le nombre de pixels dans l’image:

$$H'(c) = P(c) = \frac{\text{Nb pixels d'intensité c}}{\text{Nb total de pixels dans l'image}}$$

Ainsi défini, *P(c)* donne une idée de la probabilité d’occurrence d’un pixel de niveau de gris “c”.

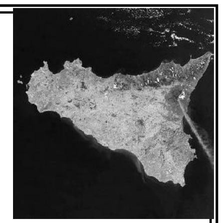
$$\sum_{c=0}^{255} P(c) = 1$$

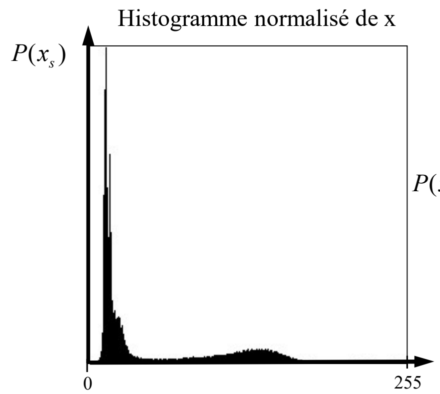
Si je tire un pixel au hasard dans l’image, j’ai 1% de chance qu’il soit d’intensité 113.



75

Quelques définitions





$P(x_s)$ Distribution des niveaux de gris dans l'image *x*.

$P(x_s = a)$ Peut se lire : probabilité d'observer un pixel de niveau de gris « a » dans l'image y.

exemple: si $P(x_s = 15) = 0.09$ alors

si je tire au hasard un pixel dans l'image x, j'aurai 9 pourcents de chance qu'il soit d'intensité 15.

Quelques définitions

$P(x_s = a)$ probabilité d'observer un pixel de niveau de gris <u>a</u> dans l'image *x*.

$P(x_s = a, mer)$ est une **probabilité jointe** qui se lit : la probabilité d'observer un pixel de niveau de gris <u>a</u> dans l'image *x* **<u>ET</u>** faisant partie de la classe mer.

$P(mer)$ Probabilité **a priori** d'observer un pixel appartenant à la classe *mer*.

$$\boxed{P(x_s = a, mer) = P(mer)P(x_s = a \mid mer)}$$

$P(x_s = a \mid mer)$ est une **probabilité conditionnelle** qui se lit : la probabilité d'observer un pixel de niveau de gris <u>a</u> dans l'image *x* **<u>ÉTANT DONNÉ</u>** qu'il appartienne à la classe mer.

Quelques définitions

$$P(x_s, y_s) = P(y_s) \times P(x_s \mid y_s)$$

terme de vraisemblance
(*likelihood* en anglais)

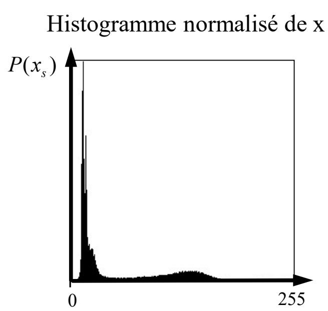
terme *a priori*

Probabilité jointe

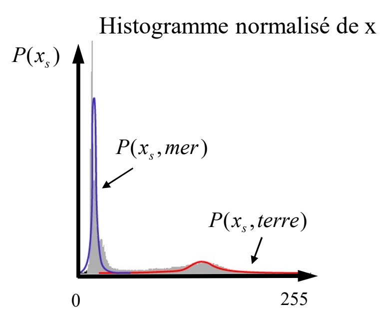
$y_s \in \{mer, terre\}$

79

Quelques définitions



$=$



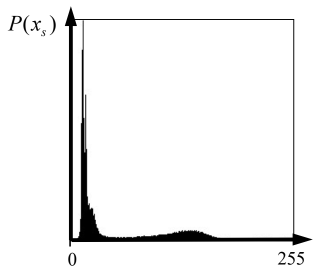
$P(x_s, terre)$ Distribution des niveaux de gris des pixels « terre »

$P(x_s, mer)$ Distribution des niveaux de gris des pixels « mer »

80

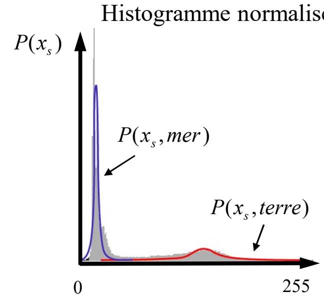
Quelques définitions

Histogramme normalisé de x



$=$

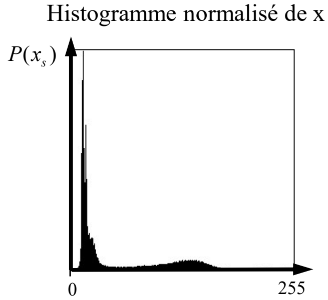
Histogramme normalisé de x



$$\boxed{P(x_s) = P(x_s, mer) + P(x_s, terre)}$$

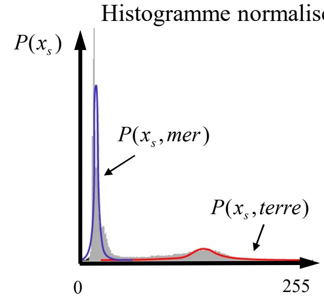
Quelques définitions

Histogramme normalisé de x



$=$

Histogramme normalisé de x



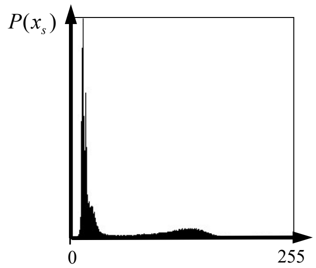
$$P(x_s) = P(x_s, mer) + P(x_s, terre)$$

Exemple

$P(x_s = 15, mer) = 0.08$ si je tire au hasard un pixel dans l'image x, j'ai 8 pourcents de chance qu'il soit d'intensité 15 **ET** qu'il appartienne à la classe **mer**

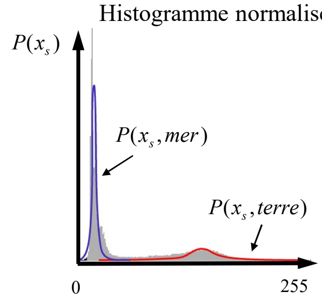
Quelques définitions

Histogramme normalisé de x



$=$

Histogramme normalisé de x



$$P(x_s) = P(x_s, mer) + P(x_s, terre)$$

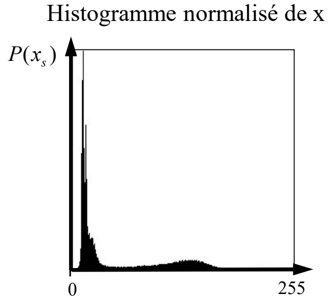
Exemple

$P(x_s = 15, terre) = 0.01$ si je tire au hasard un pixel dans l'image x, j'ai 1 pourcent de chance qu'il soit d'intensité 15 **ET** qu'il appartienne à la classe **terre**

83

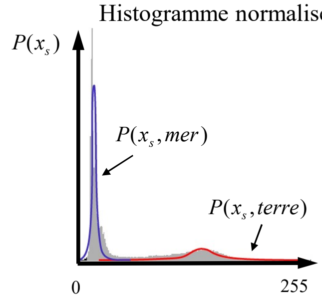
Quelques définitions

Histogramme normalisé de x



$=$

Histogramme normalisé de x



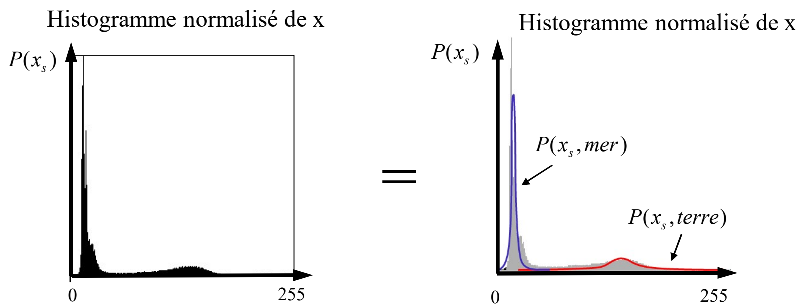
$$P(x_s) = P(x_s, mer) + P(x_s, terre)$$

Exemple $P(x_s = 15) = P(x_s = 15, mer) + P(x_s = 15, terre)$

$$0.09 \quad = \quad 0.08 \quad + \quad 0.01$$

84

Quelques définitions



$$P(x_s) = P(x_s, mer) + P(x_s, terre)$$

Exemple $P(x_s = 15) = P(x_s = 15, mer) + P(x_s = 15, terre)$

$$0.09 \quad = \quad 0.08 \quad + \quad 0.01$$

85

Quelques définitions

Ceci est un exemple de marginalisation de la variable y

$$P(x_s = 15) = P(x_s = 15, mer) + P(x_s = 15, terre)$$

$$0.09 \quad = \quad 0.08 \quad + \quad 0.01$$

$$P(x_s) = \sum_{y_s} P(x_s, y_s)$$

86

Quelques définitions

Ceci est un exemple de marginalisation de la variable x

$$P(y_s = mer) = \sum_{x_s} P(x_s, y_s = mer)$$

$$= 0.4$$

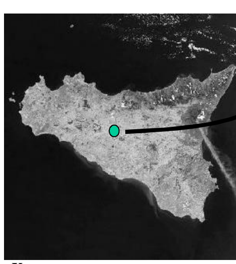
Bayes

Lorqu'on segmente une image, on cherche à déterminer pour chaque pixel, quelle classe est la plus probable. En d'autres mots, trouver l'étiquette de classe qui maximise

$$P(y_s \mid x_s)$$

Étant donné x_s on veut savoir quelle classe est la plus probable

x_s



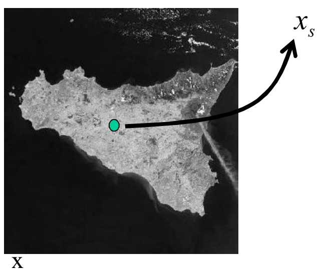
$P(y_s = terre \mid x_s)$ ou $P(y_s = mer \mid x_s)$?

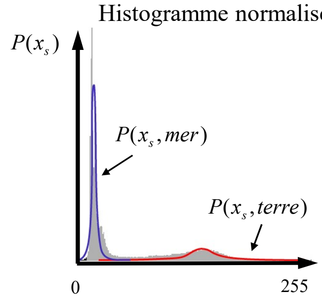
x

Bayes

Suivant la règle de Bayes

$$P(y_s \mid x_s) = \frac{P(x_s \mid y_s)P(y_s)}{P(x_s)} = \frac{P(x_s, y_s)}{P(x_s)}$$





89