RFIDEC — cours 4 : Tests d'ajustement, apprentissage de paramètres

Christophe Gonzales

LIP6 – Université Paris 6, France

Plan du cours n°4

Tests d'ajustement

Définition

- test d'ajustement = test $\Longrightarrow$ 2 issues possibles :
  1. acceptation de l'hypothèse que l'échantillon observé est tiré selon une certaine loi
  2. rejet de l'hypothèse 1
- contre-hypothèse : ne précise pas de quelle autre loi l'échantillon aurait pu être tiré

Tests d'ajustement II : le retour du χ^2

- population répartie en k classes
- échantillon de taille $n \Longrightarrow$ répartition = $(n_1, \ldots, n_k)$
- supposons l'échantillon tiré selon la loi multinomiale $(p_1, \ldots, p_k)$

  $\Longrightarrow (n_1, \ldots, n_k) \approx (n.p_1, \ldots, n.p_k)$

Rappel : $D^2_{(n)} = \sum_{r=1}^{k} \frac{(N_r - n.p_r)^2}{n.p_r} \sim \chi^2_{k-1}$

- d^2 valeur prise par $D^2_{(n)}$

  $\Longrightarrow$ si échantillon tiré selon $(p_1, \ldots, p_k)$ alors d^2 petit
- table de la loi du $\chi^2 \Longrightarrow d^2_\alpha$ tel que $P(\chi^2_{k-1} > d^2_\alpha) = \alpha$

  $\Longrightarrow$ règle de décision : si $d^2 < d^2_\alpha$ alors OK

Tests d'ajustement en pratique

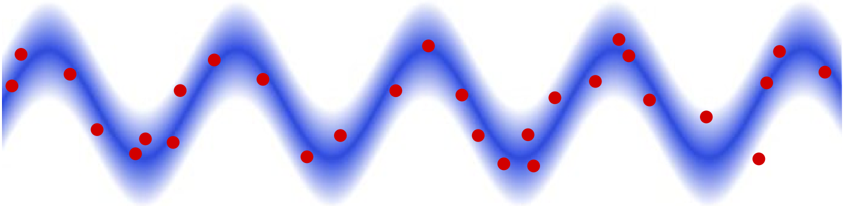
Mise en place d'un test d'ajustement

1. population répartie en k classes
2. échantillon de taille $n \Longrightarrow$ répartition $= (n_1, \ldots, n_k)$
3. on vérifie si l'échantillon tiré selon la loi $(p_1, \ldots, p_k)$:
   - A. choix du risque de première espèce α
   - B. calcul de $d^2 = \displaystyle\sum_{r=1}^{k} \frac{(n_r - n.p_r)^2}{n.p_r}$
   - C. lecture dans une table de d^2_α tel que $P(\chi^2_{k-1} > d^2_\alpha) = \alpha$
   - D. si $d^2 < d^2_\alpha$ alors règle de décision :

     $(p_1, \ldots, p_k)$ est la loi selon laquelle est tiré l'échantillon

     sinon l'échantillon est tiré selon une autre loi

Exemple de test d'ajustement (1/3)



- observations = ● = $\{(x_i, y_i)\}$
- *Problème :* les ● proviennent-ils de points situés sur la courbe $y = sin(x)$ mais observés avec un bruit gaussien ?

$\Longrightarrow$ problème : $T_i = Y_i - \sin(x_i) \sim \mathcal{N}(0, 1)$?

observations des t_i, réparties en 8 classes :

t_i	$]-\infty; -3[$	$[-3; -2[$	$[-2; -1[$	$[-1; 0[$	$[0; 1[$	$[1; 2[$	$[2; 3[$	$[3; +\infty[$
N_r	1	2	13	35	30	15	3	1

Exemple de test d'ajustement (2/3)

Rappel : $T_i \sim \mathcal{N}(0,1)$

t_i	$]-\infty;-3[$	$[-3;-2[$	$[-2;-1[$	$[-1;0[$	$[0;1[$	$[1;2[$	$[2;3[$	$[3;+\infty[$
N_r	1	2	13	35	30	15	3	1
$n.p_r$	0.14	2.14	13.59	34.13	34.13	13.59	2.14	0.14

$$\Longrightarrow d^2 = \sum_{r=1}^{8} \frac{(n_r - n.p_r)^2}{n.p_r} \approx 11.61$$

pour $\alpha = 0.05$, $P(\chi^2_7 > d^2_\alpha) = \alpha \Longrightarrow d^2_\alpha = 14.1$

$\Longrightarrow d^2 < d^2_\alpha \Longrightarrow$ règle de décision :

l'échantillon est bien tiré selon $sin(x)+$ un bruit gaussien

Exemple de test d'ajustement (3/3)

Nouvel échantillon :

t_i	$]-\infty; -3[$	$[-3; -2[$	$[-2; -1[$	$[-1; 0[$	$[0; 1[$	$[1; 2[$	$[2; 3[$	$[3; +\infty[$
N_r	2	2	12	35	30	15	3	1
$n.p_r$	0.14	2.14	13.59	34.13	34.13	13.59	2.14	0.14

$$\Longrightarrow d^2 = \sum_{r=1}^{8} \frac{(n_r - n.p_r)^2}{n.p_r} \approx 31.20$$

pour $\alpha = 0.05$, $P(\chi^2_7 > d^2_\alpha) = \alpha \Longrightarrow d^2_\alpha = 14.1$

$\Longrightarrow d^2 > d^2_\alpha \Longrightarrow$ règle de décision :

l'échantillon n'est pas tiré selon $sin(x)+$ un bruit gaussien

Exemple de test d'ajustement (1/2)

- péage d'autoroute : 10 cabines
- nombre de clients / cabine sur une heure :

N° cabine	1	2	3	4	5	6	7	8	9	10
Nb clients	24	14	18	20	23	13	23	24	23	18

Clients distribués uniformément sur l'ensemble des cabines ?

$\Longrightarrow$ test d'ajustement, niveau de confiance : $1 - \alpha = 95\%$

- H_0 = « la répartition des clients est uniforme »

  H_1 = « la répartition n'est pas uniforme »
- $H_0 \Longrightarrow$ 20 clients / cabine (uniforme)

Exemple de test d'ajustement (2/2)

- X_i : variable « effectif » recensé pour la *i*ème cabine
- Statistique d'ajustement : $D^2 = \sum_{i=1}^{10} \frac{(X_i - 20)^2}{20}$
- $D^2 \sim \chi^2_9$
- $\alpha = 0,05 = P(\text{rejeter } H_0 | H_0 \text{ est vraie})$

  $= P\left(D^2 > d_\alpha \;\middle|\; D^2 \sim \chi^2_9\right)$

  $\Longrightarrow d_\alpha = 16,9$
- calcul de la valeur de *d* observée sur l'échantillon :

$$d^2 = \frac{1}{20}[(14-20)^2 + (24-20)^2 + (18-20)^2 + (20-20)^2 + (23-20)^2 + (13-20)^2 + (23-20)^2 + (18-20)^2 + (24-20)^2 + (23-20)^2] = 7,6.$$

$\Longrightarrow$ estimation : répartition uniforme

Tests d'indépendance (1/3)

- 2 caractères X et Y
- classes de X : $A_1, A_2, \ldots, A_I$
- classes de Y : $B_1, B_2, \ldots, B_J$
- échantillon de taille n
- tableau de contingence :

$X \backslash Y$	B_1	B_2	$\cdots$	B_j	$\cdots$	B_J
A_1	n_{11}	n_{12}	$\cdots$	n_{1j}	$\cdots$	n_{1J}
A_2	n_{21}	n_{22}	$\cdots$	n_{2j}	$\cdots$	n_{2J}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
A_i	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	n_{iJ}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
A_I	n_{I1}	n_{I2}	$\cdots$	n_{Ij}	$\cdots$	n_{IJ}

Tests d'indépendance (2/3)

$X \backslash Y$	B_1	B_2	$\cdots$	B_j	$\cdots$	B_J	*total*
A_1	n_{11}	n_{12}	$\cdots$	n_{1j}	$\cdots$	n_{1J}	$n_{1.}$
A_2	n_{21}	n_{22}	$\cdots$	n_{2j}	$\cdots$	n_{2J}	$n_{2.}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
A_i	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	n_{iJ}	$n_{i.}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
A_I	n_{I1}	n_{I2}	$\cdots$	n_{Ij}	$\cdots$	n_{IJ}	$n_{I.}$
total	$n_{.1}$	$n_{.2}$	$\cdots$	$n_{.j}$	$\cdots$	$n_{.J}$	n

$$\frac{n_{ij}}{n} = P(X \in A_i, Y \in B_j)$$

$$P(X \in A_i) = \frac{n_{i.}}{n} = \frac{\sum_{j=1}^{J} n_{ij}}{n} \qquad \text{et} \qquad P(Y \in B_j) = \frac{n_{.j}}{n} = \frac{\sum_{i=1}^{I} n_{ij}}{n}$$

X et Y indépendants $\Longrightarrow P(X \in A_i, Y \in B_j) = P(X \in A_i) \times P(Y \in B_j)$

Tests d'indépendance (3/3)

$X \backslash Y$	B_1	B_2	$\cdots$	B_j	$\cdots$	B_J	*total*
A_1	n_{11}	n_{12}	$\cdots$	n_{1j}	$\cdots$	n_{1J}	$n_{1.}$
A_2	n_{21}	n_{22}	$\cdots$	n_{2j}	$\cdots$	n_{2J}	$n_{2.}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
A_i	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	n_{iJ}	$n_{i.}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
A_I	n_{I1}	n_{I2}	$\cdots$	n_{Ij}	$\cdots$	n_{IJ}	$n_{I.}$
total	$n_{.1}$	$n_{.2}$	$\cdots$	$n_{.j}$	$\cdots$	$n_{.J}$	n

X et Y indépendants $\Longrightarrow \dfrac{n_{ij}}{n} = \dfrac{n_{i.}}{n} \times \dfrac{n_{.j}}{n} \Longrightarrow n_{ij} = \dfrac{n_{i.} \times n_{.j}}{n}$

$$\chi^2_{(I-1)\times(J-1)} = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{\left(n_{ij} - \frac{n_{i.}n_{.j}}{n}\right)^2}{\frac{n_{i.}n_{.j}}{n}}$$

Exemple de test d'indépendance (1/2)

- notes d'examen de RFIDEC $\Longrightarrow$ 3 classes :

c_1	c_2	c_3
note < 8	note $\in [8, 12[$	note ≥ 12


Premier examen réparti du module RFIDEC

- X : variable aléatoire « note 1ère session »
- Y : variable aléatoire « note 2ème session »

X et Y sont-elles des variables aléatoires indépendantes ?

- sélection d'un échantillon de 100 notes :

$X \backslash Y$	c_1	c_2	c_3
c_1	2	13	6
c_2	11	27	13
c_3	3	17	8

Exemple de test d'indépendance (2/2)

Test d'indépendance de niveau de confiance 90%

1. calcul des marginales :

$X \backslash Y$	c_1	c_2	c_3	total
c_1	2	13	6	21
c_2	11	27	13	51
c_3	3	17	8	28
total	16	57	27	

2. tableau obtenu si X et Y sont indépendants :

$X \backslash Y$	c_1	c_2	c_3
c_1	3.36	11.97	5.67
c_2	8.16	29.07	13.77
c_3	4.48	15.96	7.56

3. calcul de la statistique d^2 : $d^2 = 2,42$

4. $D^2 \sim \chi^2_4 \Longrightarrow d^2_\alpha = 7,78 \Longrightarrow d^2 < d^2_\alpha \Longrightarrow$ indépendance

Rappel sur l'indépendance conditionnelle

Indépendance conditionnelle de deux variables discrètes

X et Y sont *indépendantes* conditionnellement à Z si :

- $P(X \cap Y|Z) = P(X|Z) \times P(Y|Z)$
- si $P(Y|Z) > 0$ alors $P(X|Y, Z) = P(X|Z)$
- si $P(X|Z) > 0$ alors $P(Y|X, Z) = P(Y|Z)$

Interprétation

- Conditionnement = apport de connaissances
- Si l'on connaît la valeur de la variable Z, alors connaître celle de Y n'apporte rien sur la connaissance de X

⚠ Ces formules s'étendent si X, Y et/ou Z sont remplacés par des ensembles de variables aléatoires disjoints 2 à 2

Application en informatique

- n variables aléatoires $X_1, \ldots, X_n$
- $P(X_n, \ldots, X_1) = P(X_n | X_{n-1}, \ldots, X_1) P(X_{n-1}, \ldots, X_1)$
- Par récurrence :

$$P(X_n, \ldots, X_1) = P(X_1) \times \prod_{i=2}^{n} P(X_i | X_1, \ldots, X_{i-1})$$

- $\forall i$, $\{X_1, \ldots, X_{i-1}\} = L_i \cup K_i$, où $L_i \cap K_i = \emptyset$ et X_i indépendant de L_i conditionnellement à K_i
- Alors :

$$P(X_n, \ldots, X_1) = P(X_1) \times \prod_{i=2}^{n} P(X_i | K_i)$$

- Tables de proba $P(X_i | K_i)$ plus petites que $P(X_i | X_1, \ldots, X_{i-1})$

Exemple d'application (1/5)

Exemple de la dyspnée (Lauritzen & Spiegelhalter (88))

La dyspnée peut être engendrée par une tuberculose, un cancer des poumons, une bronchite, par plusieurs de ces maladies, ou bien par aucune.
Un séjour récent en Asie augmente les chances de tuberculose, tandis que fumer augmente les risques de cancer des poumons.
Des rayons X permettent de détecter une tuberculose ou un cancer.
Un patient éprouve des difficultés à respirer. Dans quelle mesure peut-on dire qu'il est atteint de dyspnée ?

Variables aléatoires :

- D : dyspnée : oui/non
- C : cancer : oui/non
- A : Asie : oui/non
- R : rayons X : positif/négatif
- T : tuberculose : oui/non
- B : bronchite : oui/non
- F : fumer : oui/non

Exemple d'application (2/5)

Exemple de la dyspnée (Lauritzen & Spiegelhalter (88))

La dyspnée peut être engendrée par une tuberculose, un cancer des poumons, une bronchite, par plusieurs de ces maladies, ou bien par aucune. Un séjour récent en Asie augmente les chances de tuberculose, tandis que fumer augmente les risques de cancer des poumons. Des rayons X permettent de détecter une tuberculose ou un cancer. Un patient éprouve des difficultés à respirer. Dans quelle mesure peut-on dire qu'il est atteint de dyspnée ?

$$P(D, R, T, C, B, A, F) = P(D|R, T, C, B, A, F) \times P(R, T, C, B, A, F)$$

Or $P(D|R, T, C, B, A, F) = P(D|T, C, B)$

$$\Longrightarrow P(D, R, T, C, B, A, F) = P(D|T, C, B) \times P(R, T, C, B, A, F)$$

Exemple d'application (3/5)

Exemple de la dyspnée (Lauritzen & Spiegelhalter (88))

La dyspnée peut être engendrée par une tuberculose, un cancer des poumons, une bronchite, par plusieurs de ces maladies, ou bien par aucune. Un séjour récent en Asie augmente les chances de tuberculose, tandis que fumer augmente les risques de cancer des poumons. Des rayons X permettent de détecter une tuberculose ou un cancer. Un patient éprouve des difficultés à respirer. Dans quelle mesure peut-on dire qu'il est atteint de dyspnée ?

$P(D, R, T, C, B, A, F) = P(D|T, C, B) \times P(R, T, C, B, A, F)$

or $P(R, T, C, B, A, F) = P(R|T, C, B, A, F) \times P(T, C, B, A, F)$

et $P(R|T, C, B, A, F) = P(R|T, C)$

$\Longrightarrow P(D, R, T, C, B, A, F) = P(D|T, C, B) \times P(R|T, C) \times P(T, C, B, A, F)$

Exemple d'application (4/5)

Exemple de la dyspnée (Lauritzen & Spiegelhalter (88))

La dyspnée peut être engendrée par une tuberculose, un cancer des poumons, une bronchite, par plusieurs de ces maladies, ou bien par aucune. Un séjour récent en Asie augmente les chances de tuberculose, tandis que fumer augmente les risques de cancer des poumons. Des rayons X permettent de détecter une tuberculose ou un cancer. Un patient éprouve des difficultés à respirer. Dans quelle mesure peut-on dire qu'il est atteint de dyspnée ?

$$P(D,R,T,C,B,A,F) = P(D|T,C,B) \times P(R|T,C) \times P(T,C,B,A,F)$$

$$P(T|C,B,A,F) = P(T|A)$$

$$P(D,R,T,C,B,A,F) = P(D|T,C,B) \times P(R|T,C) \times P(T|A) \times P(C,B,A,F)$$

....

$$= P(D|T,C,B) \times P(R|T,C) \times P(T|A) \times P(C|F) \times P(B|F) \times P(A) \times P(F)$$

Exemple d'application (5/5)

$$P(D, R, T, C, B, A, F) = P(D|T, C, B) \times P(R|T, C) \times P(T|A) \times P(C|F) \times P(B|F) \times P(A) \times P(F)$$

Si toutes les variables ont 10 valeurs possibles :

$P(D, R, T, C, B, A, F)$ nécessite une table de 10^7 éléments

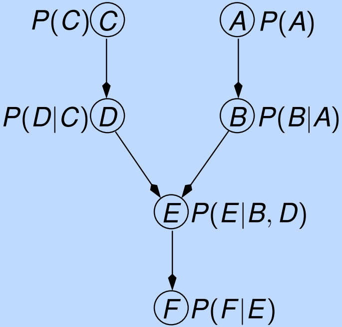
formule décomposée nécessite :

$10000 + 1000 + 3 \times 100 + 2 \times 10 = 11320$ éléments

Qu'est-ce qu'un réseau bayésien ?

Définition d'un réseau bayésien

1. un graphe sans circuit :



   qui représente une décomposition de la loi jointe :
   $P(A,B,C,D,E,F) = P(F|E)P(E|B,D)P(D|C)P(C)P(B|A)P(A)$

2. À chaque noeud X du graphe est associé sa probabilité conditionnellement à ses parents.

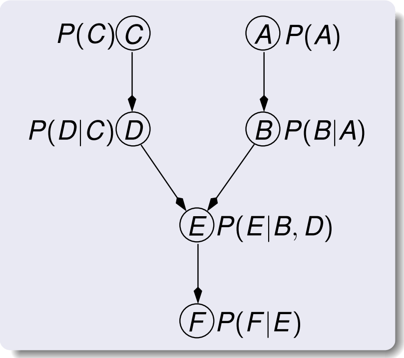
Quelques applications des réseaux bayésiens

diagnostic

- *diagnostic de panne*
- *sûreté de fonctionnement*
- *filtrage de spams*

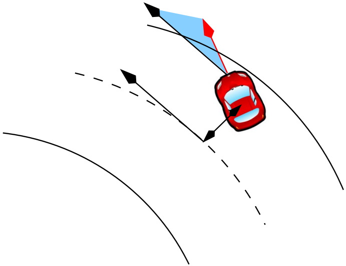
prédiction

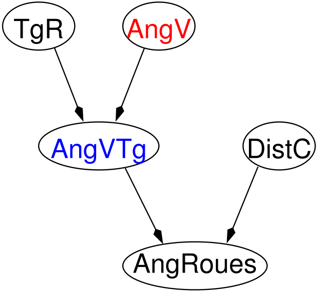
- *modélisation de joueurs*
- *prévisions boursières*



Diagnostic : une voiture bayésienne (1/3)

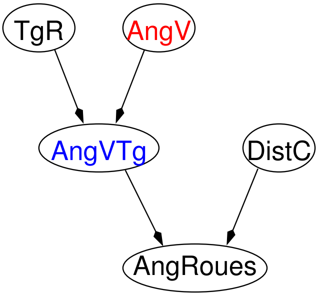
- RB conduit une voiture
- but : rester au milieu de la route





Problème : effectuer rapidement les calculs

Diagnostic : une voiture bayésienne (2/3)



décomposition :

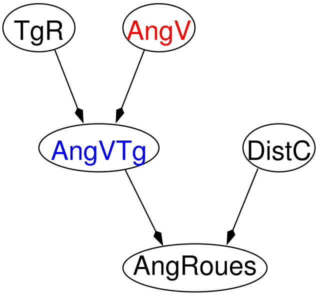
P(AngRoues|AngVTg,DistC)

P(AngVTg|TgR,AngV)

P(TgR)P(AngV)P(DistC)

Principe de fonctionnement de la voiture

- calculer P(AngRoues) puis tirer selon cette loi une action
- calculer Argmax P(AngRoues)
- calculs $\Longrightarrow$ éliminer les variables une par une



décomposition :

$P(\text{AngRoues}|\text{AngVTg,DistC})$

$P(\text{AngVTg}|\text{TgR,AngV})$

$P(\text{TgR})P(\text{AngV})P(\text{DistC})$

Vraisemblance d'un échantillon : le cas discret

- Paramètre à estimer : Θ
- Échantillon $\mathbf{x} = (x_1, \dots, x_n)$ de taille n
- Échantillon $\Longrightarrow$ les x_i = réalisations de variables aléatoires X_i
- Échantillon i.i.d. $\Longrightarrow$ les X_i sont mutuellement indépendants

$$\Longrightarrow P(X_1 = x_1, \dots, X_n = x_n | \Theta = \theta) = \prod_{i=1}^{n} P(X_i = x_i | \Theta = \theta)$$

Vraisemblance d'un échantillon dans le cas discret

- $L(\mathbf{x}, \theta)$ = Vraisemblance de l'échantillon
- $L(\mathbf{x}, \theta)$ = proba d'obtenir cet échantillon sachant que $\Theta = \theta$

$$L(\mathbf{x}, \theta) = P(x_1, \dots, x_n | \Theta = \theta) = \prod_{i=1}^{n} P(x_i | \Theta = \theta)$$

- Paramètre à estimer : Θ
- Échantillon $\mathbf{x} = (x_1, \ldots, x_n)$ de taille n
- Échantillon i.i.d. $\Longrightarrow$ les X_i sont mutuellement indépendants
- p : fonction de densité

$$\Longrightarrow p(X_1 = x_1, \ldots, X_n = x_n | \Theta = \theta) = \prod_{i=1}^{n} p(X_i = x_i | \Theta = \theta)$$

Vraisemblance d'un échantillon dans le cas continu

- $L(\mathbf{x}, \theta) =$ Vraisemblance de l'échantillon
- $L(\mathbf{x}, \theta) = p(x_1, \ldots, x_n | \Theta = \theta) = \prod_{i=1}^{n} p(x_i | \Theta = \theta)$

Exemple de vraisemblance (1/2)

- pièce de monnaie : $P(\text{Pile}) = 0,75$ et $P(\text{Face}) = 0,25$
- jet de la pièce $\Longrightarrow$ expérience de Bernoulli
- paramètre θ = proba de Pile $= 0,75 = \theta$

- échantillon 1 :

P	P	F	F	P	P	F	P	P	P

$$\Longrightarrow L(\mathbf{x}, \theta) = \prod_{i=1}^{7} P(\text{Pile}|\theta) \times \prod_{i=1}^{3} P(\text{Face}|\theta)$$

$$= 0,75^7 \times 0,25^3 \approx 0,002086$$

- échantillon 2 :

F	F	P	P	F	F	P	F	F	F

$$\Longrightarrow L(\mathbf{x}, \theta) = \prod_{i=1}^{3} P(\text{Pile}|\theta) \times \prod_{i=1}^{7} P(\text{Face}|\theta)$$

$$= 0,75^3 \times 0,25^7 \approx 0,000026$$

Exemple de vraisemblance (2/2)

- pièce de monnaie : $P(\text{Pile}) = ???$ et $P(\text{Face}) = ???$
- paramètre θ = proba de Pile = ???
- échantillon :

P	P	F	F	P	P	F	P	P	P

$$\Longrightarrow L(\mathbf{x}, \theta) = \prod_{i=1}^{7} P(\text{Pile}|\theta) \times \prod_{i=1}^{3} P(\text{Face}|\theta)$$

- $\theta_1 = 0,75 \Longrightarrow L(x, \theta_1) = 0,75^7 \times 0,25^3 \approx 0,002086$
- $\theta_2 = 0,5 \Longrightarrow L(x, \theta_2) = 0,5^7 \times 0,5^3 \approx 0,000976$
- $\theta_3 = 0,25 \Longrightarrow L(x, \theta_3) = 0,25^7 \times 0,75^3 \approx 0,000026$

$\Longrightarrow \theta_1$ plus vraisemblable que θ_2 ou θ_3

Estimateur du maximum de vraisemblance

Estimateur du maximum de vraisemblance

- X : variable aléatoire sur la population
- X suit une loi de proba de paramètre θ inconnu
- Θ : ensemble des valeurs possibles pour θ
- $\mathbf{x}$: échantillon i.i.d.
- $T = f(X) =$ *estimateur du maximum de vraisemblance*

  défini par $\mathbf{x} \longmapsto t = f(\mathbf{x}) = \underset{\theta \in \Theta}{\text{Argmax}}\, L(\mathbf{x}, \theta)$

$\Longrightarrow t =$ valeur θ de Θ pour laquelle la proba d'observer $\mathbf{x}$ était la plus grande

Calcul du maximum de vraisemblance

Problème : comment calculer le maximum de vraisemblance ?

- $\underset{\theta \in \Theta}{\text{Argmax}}\, L(\mathbf{x}, \theta) = \underset{\theta \in \Theta}{\text{Argmax}}\, P(x_1, \ldots, x_n | \theta) = \underset{\theta \in \Theta}{\text{Argmax}} \prod_{i=1}^{n} P(x_i | \theta)$
- Certaines conditions de concavité et de dérivabilité

$$\Longrightarrow \underset{\theta \in \Theta}{\text{Argmax}}\, L(\mathbf{x}, \theta) \text{ obtenu lorsque } \frac{\partial L(\mathbf{x}, \theta)}{\partial \theta} = 0$$

- $\underset{\theta \in \Theta}{\text{Argmax}}\, L(\mathbf{x}, \theta) = \underset{\theta \in \Theta}{\text{Argmax}} \ln L(\mathbf{x}, \theta) = \underset{\theta \in \Theta}{\text{Argmax}} \sum_{i=1}^{n} \ln P(x_i | \theta)$

$$\underset{\theta \in \Theta}{\text{Argmax}} \ln L(\mathbf{x}, \theta) = \text{log vraisemblance}$$

$$\Longrightarrow \underset{\theta \in \Theta}{\text{Argmax}}\, L(\mathbf{x}, \theta) \text{ obtenu lorsque } \sum_{i=1}^{n} \frac{\partial \ln P(x_i | \theta)}{\partial \theta} = 0$$

Max de vraisemblance et loi normale (1/2)

- $X \sim \mathcal{N}(\mu, \sigma^2)$; on suppose $\sigma = 1$
- paramètre θ = espérance μ
- loi normale $\Longrightarrow$ vraisemblance :

$$L(\mathbf{x}, \theta) = \prod_{i=1}^{n} p(x_i|\theta) = \prod_{i=1}^{n} \left[\frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}(x_i - \theta)^2 \right\} \right]$$

- $\dfrac{\partial L(\mathbf{x}, \theta)}{\partial \theta} = 0 \Longleftrightarrow \dfrac{\partial \ln L(\mathbf{x}, \theta)}{\partial \theta} = 0$
- $\ln L(\mathbf{x}, \theta) = -\dfrac{n}{2} \ln 2\pi - \dfrac{1}{2} \displaystyle\sum_{i=1}^{n} (x_i - \theta)^2$
- $\dfrac{\partial L(\mathbf{x}, \theta)}{\partial \theta} = 0 \Longleftrightarrow \displaystyle\sum_{i=1}^{n} (x_i - \theta) = 0 \Longleftrightarrow \theta = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} x_i = \overline{x}$

Estimateur du maximum de vraisemblance : $\overline{X}$

Max de vraisemblance et loi normale (2/2)

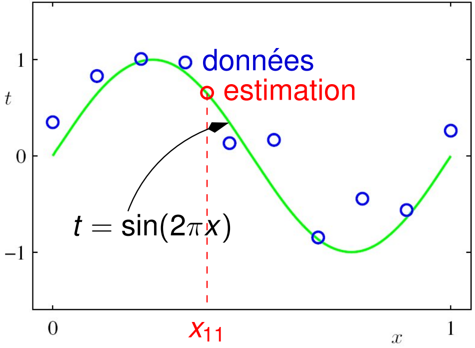
- $X \sim \mathcal{N}(\mu, \sigma^2)$
- paramètre $\theta = (\mu, \sigma^2)$
- Log vraisemblance :

$$\ln L(\mathbf{x}, \theta) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$$

- Maximum de vraisemblance $\Longrightarrow \dfrac{\partial L(\mathbf{x}, \theta)}{\partial \mu} = 0$ et $\dfrac{\partial L(\mathbf{x}, \theta)}{\partial \sigma^2} = 0$
- $$\begin{cases} \dfrac{\partial L(\mathbf{x}, \theta)}{\partial \mu} = \dfrac{1}{\sigma^2} \displaystyle\sum_{i=1}^{n} (x_i - \mu) = 0 & \Longrightarrow \mu = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} x_i = \overline{x} \\ \dfrac{\partial L(\mathbf{x}, \theta)}{\partial \sigma^2} = -\dfrac{n}{2} \dfrac{1}{\sigma^2} + \dfrac{1}{2\sigma^4} \displaystyle\sum_{i=1}^{n} (x_i - \mu)^2 = 0 & \Longrightarrow \sigma^2 = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} (x_i - \overline{x})^2 = s_n^2 \end{cases}$$

Estimateurs du maximum de vraisemblance : $\overline{X}$ et S_n^2

⚠ estimateur de la variance biaisé : variance non corrigée



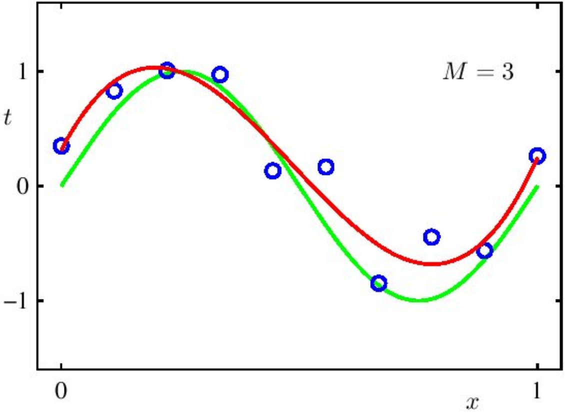
Observations
(x_1, t_1)
$\vdots$
(x_{10}, t_{10})

$\Longrightarrow$ courbe $\sin(2\pi x)$ $\Longrightarrow$ estimation de t_{11}

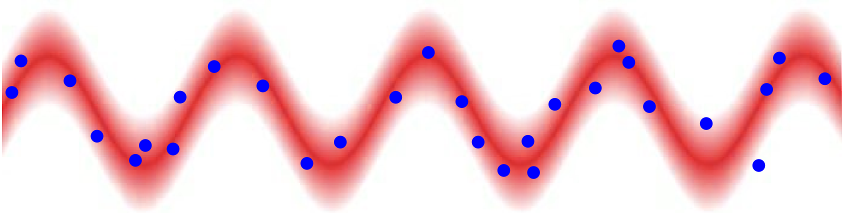
$\Longrightarrow$ reconnaissance de la courbe verte

Idée : estimer la courbe verte par un polynôme :

$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \cdots + w_M x^M = \sum_{j=0}^{M} w_j x^j$$



Idée : les ordonnées des points bleus sont distribuées selon une loi normale autour de $y(x, \mathbf{w})$:



$\Longrightarrow P(t|x, \mathbf{w}, \sigma^2) = \mathcal{N}(t|y(x, \mathbf{w}), \sigma^2)$

Problème : comment trouver $\mathbf{w}$ et σ^2 ?

$\Longrightarrow$ par maximum de vraisemblance

$$P(t|x, \mathbf{w}, \sigma^2) = \mathcal{N}(t|y(x, \mathbf{w}), \sigma^2)$$

- observations $\{(x_i, t_i), i = 1, \ldots, n\}$
- $\mathbf{t} = \{t_1, \ldots, t_n\}$; $\quad \mathbf{x} = \{x_1, \ldots, x_n\}$
- observations $\Longrightarrow$ échantillon i.i.d

$$\Longrightarrow P(\mathbf{t}|\mathbf{x}, \mathbf{w}, \sigma^2) = \prod_{i=1}^{n} P(t_i|x_i, \mathbf{w}, \sigma^2)$$

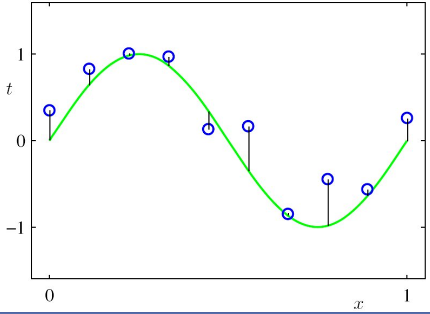
$$= \prod_{i=1}^{n} \mathcal{N}(t_i|y(x_i, \mathbf{w}), \sigma^2)$$

- Max de vraisemblance $\Longrightarrow$ calculer la log-vraisemblance :

$$\ln p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \sigma^2) = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} [y(x_i, \mathbf{w}) - t_i]^2 + \frac{n}{2} \ln \frac{1}{\sigma^2} - \frac{n}{2} \ln(2\pi)$$

$$\ln p(\mathbf{t}|\mathbf{x},\mathbf{w},\sigma^2) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}[y(x_i,\mathbf{w}) - t_i]^2 + \frac{n}{2}\ln\frac{1}{\sigma^2} - \frac{n}{2}\ln(2\pi)$$

- Maximum de log-vraisemblance $\Longrightarrow$ trouver $\mathbf{w}_{ML}$ et σ^2_{ML} qui maximisent $\ln p(\mathbf{t}|\mathbf{x},\mathbf{w},\sigma^2)$
- maximiser par rapport à $\mathbf{w}_{ML} \Longleftrightarrow$ minimiser $\sum_{i=1}^{n}[y(x_i,\mathbf{w}) - t_i]^2$



$\Longrightarrow$ critère du cours 1.1

Retour sur le problème d'ajustement du cours 1 (6/6)

$$\ln p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \sigma^2) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}[y(x_i, \mathbf{w}) - t_i]^2 + \frac{n}{2}\ln\frac{1}{\sigma^2} - \frac{n}{2}\ln(2\pi)$$

- maximiser $\ln p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \sigma^2)$ par rapport à $\sigma^2 \Longrightarrow \dfrac{\partial \ln p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \sigma^2)}{\partial \sigma^2} = 0$
- $\dfrac{\partial \ln p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \sigma^2)}{\partial \sigma^2} = \dfrac{1}{2\sigma^4}\displaystyle\sum_{i=1}^{n}[y(x_i, \mathbf{w}) - t_i]^2 - \dfrac{n}{2\sigma^4}\sigma^2 = 0$

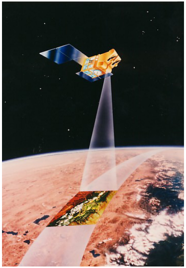
$$\Longrightarrow \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}[y(x_i, \mathbf{w}) - t_i]^2$$

$$\sigma^2_{ML} = \frac{1}{n}\sum_{i=1}^{n}[y(x_i, \mathbf{w}_{ML}) - t_i]^2$$

Prévention des risques d'inondation (1/4)

- Plan de prévention des risques d'inondations (PPR-I) :

photos satellite SPOT5 $\Longrightarrow$ zones susceptibles d'être inondées



- 3 catégories de parcelles :
  1. inondables (*PI*)
  2. partiellement inondables (*PPI*)
  3. non inondables (*NI*)

Prévention des risques d'inondation (2/4)

- images en teintes de gris
- proba d'obtenir un niveau de gris n dépend du type de zone :

$$P(n|PI) = \mathcal{N}(\mu_1, \sigma_1^2) \qquad P(n|PPI) = \mathcal{N}(\mu_2, \sigma_2^2)$$
$$\mu_1 = 100 \quad \sigma_1 = 20 \qquad \mu_2 = 85 \quad \sigma_2 = 5$$

- nouvelle image envoyée par SPOT5 :

- zone Z : niveau de gris $= n = 80$

Problème : zone $Z = PI$ ou PPI ?

Problème : zone $Z = PI$ ou PPI ?

- 2 hypothèses :
  1. $\theta_1 = \ll Z$ est de type $PI \gg$
  2. $\theta_2 = \ll Z$ est de type $PPI \gg$
- Idée : calcul du max de vraisemblance d'obtenir la zone Z sous θ_1 ou sous θ_2
- $L(\mathbf{x}, \theta_1) = p(80|PI)$, avec p fct de densité de $P(n|PI) = \mathcal{N}(\mu_1, \sigma_1^2)$

Rappel : la fonction de densité de $\mathcal{N}(\mu, \sigma^2)$ est :

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

Prévention des risques d'inondation (4/4)

Problème : zone $Z = PI$ ou PPI ?

- $P(n|PI) = \mathcal{N}(\mu_1, \sigma_1^2) = \mathcal{N}(100, 20^2)$
- $L(\mathbf{x}, \theta_1) = p(80|PI)$

$$= \frac{1}{\sqrt{2\pi} \times 20} \exp\left\{-\frac{1}{2}\left(\frac{80-100}{20}\right)^2\right\}$$

$$= \frac{1}{20\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\right\}$$

$$\approx 0,0121$$

- $P(n|PPI) = \mathcal{N}(\mu_2, \sigma_2^2) = \mathcal{N}(85, 5^2)$
- $L(\mathbf{x}, \theta_2) = p(80|PPI) = \frac{1}{\sqrt{2\pi} \times 5} \exp\left\{-\frac{1}{2}\left(\frac{80-85}{5}\right)^2\right\} \approx 0,0484$

Max de vraisemblance $\Longrightarrow$ *PPI* plus probable