RFIDEC — cours 5: MAP et apprentissage non paramétrique

Christophe Gonzales

LIP6 – Université Paris 6, France

1. Maximum a posteriori
2. Estimation de densité

Max de vraisemblance et loi binomiale (1/2)

- pièce de monnaie
- $X \in \{0, 1\}$, $0 \iff$ Face, $1 \iff$ Pile
- $X \sim \mathcal{B}(1, p) \Longrightarrow P(X = x|p) = p^x(1 - p)^{1-x}$
- n lancers de la pièce $\Longrightarrow$ observations $\mathbf{x} = \{x_1, \ldots, x_n\}$
- $P(\mathbf{x}|p) = \displaystyle\prod_{i=1}^{n} P(x_i|p) = \prod_{i=1}^{n} p^{x_i}(1 - p)^{1-x_i}$

Problème : à partir de **x**, peut-on raisonnablement déduire p ?

- maximum de vraisemblance :

$$\ln P(\mathbf{x}|p) = \sum_{i=1}^{n} [x_i \ln p + (1 - x_i) \ln(1 - p)]$$

$$\frac{\partial \ln P(\mathbf{x}|p)}{\partial p} = \frac{1}{p}\sum_{i=1}^{n} x_i - \frac{n - \sum_{i=1}^{n} x_i}{1 - p} = 0 \Longrightarrow p_{ML} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Max de vraisemblance et loi binomiale (2/2)

$$p_{ML} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- 3 lancers $\Longrightarrow$ observations : {Pile,Pile,Pile}

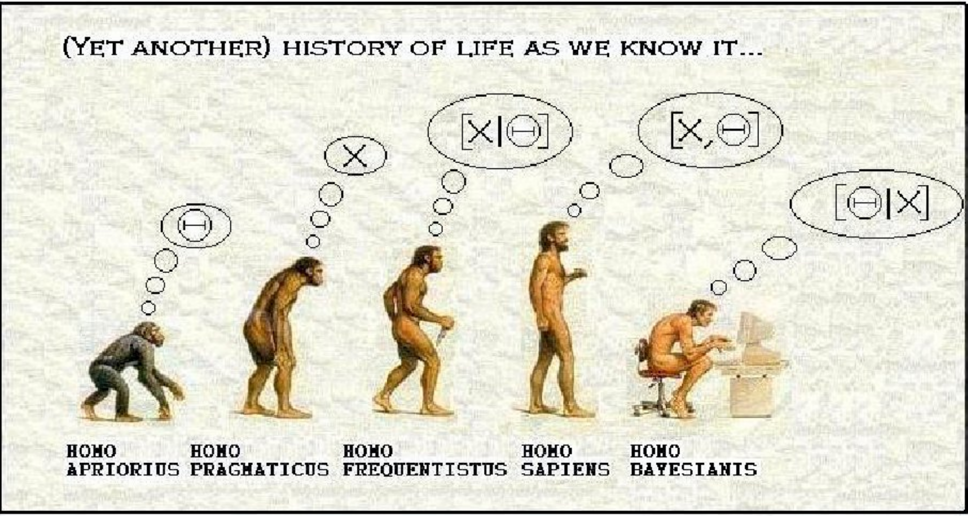
- Maximum de vraisemblance $\Longrightarrow p_{ML} = 1$

$\Longrightarrow$ on considère que tout lancer de la pièce devrait tomber sur Pile

$\Longrightarrow$ résultat à l'encontre du bon sens

$\Longrightarrow$ autre estimateur : maximum a posteriori

En route vers le maximum a posteriori



Le modèle bayésien (1/4)

Maximum a posteriori $\Longrightarrow$ modèle bayésien

Modèle bayésien

événements : parties de $\mathcal{X} \times \Theta$, où :

- $\mathcal{X} =$ l'espace des observations (échantillons) **x** de taille *n*
- $\Theta =$ espace des paramètres θ
- famille des événements dotée d'une loi de proba Π

- *cas discret :* Π déterminée par les probas des événements élémentaires $\pi(\mathbf{x}, \theta)$
- *cas continu :* Π déterminée par la densité jointe $\pi(\mathbf{x}, \theta)$

⚠ Max de vraisemblance : $\pi(\mathbf{x}|\theta)$ au lieu de $\pi(\mathbf{x}, \theta) = \pi(\mathbf{x}|\theta)\pi(\theta)$

Le modèle bayésien (2/4)

Le cas discret :

- $\pi(\mathbf{x}, \theta) = \Pi(X = \mathbf{x}, \Theta = \theta)$, où X, Θ variables aléatoires
- $\pi(\mathbf{x}) = \Pi(X = \mathbf{x}) = \sum_{\theta \in \Theta} \Pi(X = \mathbf{x}, \Theta = \theta) = \sum_{\theta \in \Theta} \pi(\mathbf{x}, \theta)$
- $\pi(\theta) = \Pi(\Theta = \theta) = \sum_{\mathbf{x} \in \mathcal{X}} \Pi(X = \mathbf{x}, \Theta = \theta) = \sum_{\mathbf{x} \in \mathcal{X}} \pi(\mathbf{x}, \theta)$
- $\pi(\mathbf{x}|\theta) = \Pi(X = \mathbf{x}|\Theta = \theta) = \dfrac{\pi(\mathbf{x}, \theta)}{\pi(\theta)}$
- $\pi(\theta|\mathbf{x}) = \Pi(\Theta = \theta|X = \mathbf{x}) = \dfrac{\pi(\mathbf{x}, \theta)}{\pi(\mathbf{x})}$

Probabilités a priori et a posteriori

- $\pi(\theta)$ = probabilité a priori de θ
- $\pi(\theta|\mathbf{x})$ = probabilité a posteriori de θ

Le cas continu :

- $\pi(\mathbf{x}) = \Pi(X = \mathbf{x}) = \int_{\theta \in \Theta} \Pi(X = \mathbf{x}, \Theta = \theta) d\theta = \int_{\theta \in \Theta} \pi(\mathbf{x}, \theta) d\theta$
- $\pi(\theta) = \Pi(\Theta = \theta) = \int_{\mathbf{x} \in \mathcal{X}} \Pi(X = \mathbf{x}, \Theta = \theta) d\mathbf{x} = \int_{\mathbf{x} \in \mathcal{X}} \pi(\mathbf{x}, \theta) d\mathbf{x}$
- $\pi(\mathbf{x}|\theta) = \dfrac{\pi(\mathbf{x}, \theta)}{\pi(\theta)}$
- $\pi(\theta|\mathbf{x}) = \dfrac{\pi(\mathbf{x}, \theta)}{\pi(\mathbf{x})}$

Probabilités a priori et a posteriori

- $\pi(\theta) = $ probabilité a priori de θ
  $= $ idée que l'on se fait de θ avant observation
- $\pi(\theta|\mathbf{x}) = $ probabilité a posteriori de θ
  $= $ idée que l'on se fait de θ après observation

- Formule de Bayes : $\pi(\theta|\mathbf{x}) = \dfrac{\pi(\mathbf{x}|\theta)\pi(\theta)}{\pi(\mathbf{x})}$

- $$\begin{cases} \textit{cas discret :}\ \pi(\theta|\mathbf{x}) = \dfrac{\pi(\mathbf{x}|\theta)\pi(\theta)}{\sum_{\theta\in\Theta}\pi(\mathbf{x},\theta)} = \dfrac{\pi(\mathbf{x}|\theta)\pi(\theta)}{\sum_{\theta\in\Theta}\pi(\mathbf{x}|\theta)\pi(\theta)} \\ \textit{cas continu :}\ \pi(\theta|\mathbf{x}) = \dfrac{\pi(\mathbf{x}|\theta)\pi(\theta)}{\int_{\theta\in\Theta}\pi(\mathbf{x},\theta)d\theta} = \dfrac{\pi(\mathbf{x}|\theta)\pi(\theta)}{\int_{\theta\in\Theta}\pi(\mathbf{x}|\theta)\pi(\theta)d\theta} \end{cases}$$

- *Rappel :* $\pi(\mathbf{x}|\theta) = $ vraisemblance de l'échantillon $= L(\mathbf{x},\theta)$

Maximum a posteriori

Maximum a posteriori (MAP)

T estimateur du maximum a posteriori de θ :

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

- échantillon i.i.d de n observations
- $X = (X_1, \ldots, X_n) \Longrightarrow \mathbf{x} = (x_1, \ldots, x_n)$ observation de X
- $$\begin{cases} \textit{cas discret :}\ \pi(\theta|\mathbf{x}) = \dfrac{L(\mathbf{x},\theta)\pi(\theta)}{\sum_{\theta\in\Theta} L(\mathbf{x},\theta)\pi(\theta)} \\ \textit{cas continu :}\ \pi(\theta|\mathbf{x}) = \dfrac{L(\mathbf{x},\theta)\pi(\theta)}{\int_{\theta\in\Theta} L(\mathbf{x},\theta)\pi(\theta)d\theta} \end{cases}$$
- échantillon i.i.d $\Longrightarrow \pi(\mathbf{x}|\theta) = L(\mathbf{x},\theta) = \begin{cases} \prod_{i=1}^n P(x_i|\theta) & \text{(discret)} \\ \prod_{i=1}^n p(x_i|\theta) & \text{(continu)} \end{cases}$

- pièce de monnaie $\Longrightarrow X \in \{0, 1\}$

  $0 \Longleftrightarrow$ Face  $1 \Longleftrightarrow$ Pile 

- $X \sim \mathcal{B}(1, \theta) \Longrightarrow P(X = x|\theta) = \theta^x(1-\theta)^{1-x}$
- échantillon **x** de 3 lancers $\Longrightarrow$ {Pile,Pile,Pile} 
- Max de vraisemblance $\Longrightarrow \theta_{ML} = 1$

  $\Longrightarrow$ tous les lancers devraient tomber sur Pile

---

- Modèle bayésien : $\Theta = \{\theta_1, \theta_2\}$

  $\theta_1 =$ « biais en faveur de Pile », $\theta_2 =$ « biais en faveur de Face »

- Info a priori : $\pi(\theta_1) = \frac{2}{3}$, $\pi(\theta_2) = \frac{1}{3}$

Problème : quelle est la valeur du maximum a posteriori ?

MAP : retour sur la pièce de monnaie (2/2)

- Info a priori : $\pi(\theta_1) = \frac{2}{3}$, $\pi(\theta_2) = \frac{1}{3}$
- $L(\mathbf{x}, \theta_1) = \pi(\mathbf{x}|\theta_1) = \prod_{i=1}^{3} P(x_i|\theta_1) = \frac{2}{3}^3 \times \left(1 - \frac{2}{3}\right)^0 = \frac{2}{3}^3 \approx 0,296$
- $L(\mathbf{x}, \theta_2) = \pi(\mathbf{x}|\theta_2) = \prod_{i=1}^{3} P(x_i|\theta_2) = \frac{1}{3}^3 \times \left(1 - \frac{1}{3}\right)^0 = \frac{1}{3}^3 \approx 0,037$
- $\pi(\theta_1|\mathbf{x}) = \dfrac{L(\mathbf{x}, \theta_1)\pi(\theta_1)}{\sum_{\theta\in\Theta} L(\mathbf{x}, \theta)\pi(\theta)} = \dfrac{\frac{2}{3}^3 \times \frac{2}{3}}{\frac{2}{3}^3 \times \frac{2}{3} + \frac{1}{3}^3 \times \frac{1}{3}} \approx 0,941$
- $\pi(\theta_2|\mathbf{x}) = \dfrac{L(\mathbf{x}, \theta_2)\pi(\theta_2)}{\sum_{\theta\in\Theta} L(\mathbf{x}, \theta)\pi(\theta)} \approx 0,059$

Max a posteriori : $\theta = \theta_1 \Longrightarrow X \sim \mathcal{B}(1, \theta_1) = \mathcal{B}(1, 0.941)$

⚠ probabilité que la pièce tombe sur Face $\neq 0$

MAP et les lois conjuguées

calcul de la distribution a posteriori : $\pi(\theta|\mathbf{x}) = \dfrac{\pi(\mathbf{x}|\theta)\pi(\theta)}{\int_{\theta\in\Theta}\pi(\mathbf{x}|\theta)\pi(\theta)d\theta}$

$\Longrightarrow$ si $\pi(\mathbf{x}|\theta)\pi(\theta)$ complexe analytiquement alors calcul de l'intégrale compliqué

Lois conjuguées

- $\pi(\theta)$: loi a priori
- $\pi(\mathbf{x}|\theta)$: fonction de vraisemblance
- $\pi(\theta|\mathbf{x})$: distribution a posteriori
- $\pi(\theta)$ et $\pi(\mathbf{x}|\theta)$ sont conjuguées si $\pi(\theta|\mathbf{x})$ appartient à la même famille de lois que $\pi(\theta)$

Lois conjuguées : exemple de la pièce de monnaie

- pièce de monnaie $\Longrightarrow X \in \{0, 1\}$: $0 \Longleftrightarrow$  $1 \Longleftrightarrow$ 
- $X \sim \mathcal{B}(1, \theta) \Longrightarrow$ vraisemblance d'un échantillon :

  $\pi(\mathbf{x}|\theta) = \theta^x(1-\theta)^{n-x}$, avec $x = \#(x_i = 1)$

  $\Longrightarrow$ loi binomiale

Distribution de probabilité Beta

Loi Beta : $\text{Beta}(\theta, a, b) = \dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$

avec $\Gamma(x) = \displaystyle\int_0^{+\infty} t^{x-1}e^{-t}dt$

Espérance $= \dfrac{a}{a+b}$ Variance $= \dfrac{ab}{(a+b)^2(a+b+1)}$

$\Longrightarrow$ loi Beta et loi binomiales conjuguées

Lois conjuguées : loi binomiale et loi Beta

- loi a priori : $\pi(\theta) = \text{Beta}(\theta, a, b) = \dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$
- fonction de vraisemblance : $\pi(\mathbf{x}|\theta) = \theta^x(1-\theta)^{n-x}$, avec $x = \#(x_i = 1)$
- loi a posteriori : $\pi(\theta|\mathbf{x}) = \dfrac{\pi(\mathbf{x}|\theta)\pi(\theta)}{\sum_{\theta\in\Theta}\pi(\mathbf{x}|\theta)\pi(\theta)} \propto \pi(\mathbf{x}|\theta)\pi(\theta)$
- loi a posteriori : $\pi(\theta|\mathbf{x}) \propto \theta^{x+a-1}(1-\theta)^{b+n-x-1}$

$$\Longrightarrow \pi(\theta|\mathbf{x}) \sim \text{Beta}(\theta, x+a, b+n-x)$$

Comparaison MAP – maximum de vraisemblance

- pièce de monnaie $\Longrightarrow X \in \{0,1\}$: $0 \Longleftrightarrow$  $1 \Longleftrightarrow$ 
- Max de vraisemblance :

$$\pi(\mathbf{x}|\theta) = \theta^x(1-\theta)^{n-x} \Longrightarrow \text{Beta}(\theta, x+1, n-x+1)$$

- Max a posteriori :

$$\pi(\theta|\mathbf{x}) \propto \theta^{x+a-1}(1-\theta)^{b+n-x-1} \Longrightarrow \text{Beta}(\theta, x+a, n-x+b)$$

$\Longrightarrow$ Max de vraisemblance $\Longleftrightarrow$ Max a posteriori avec $a = 1$ et $b = 1$

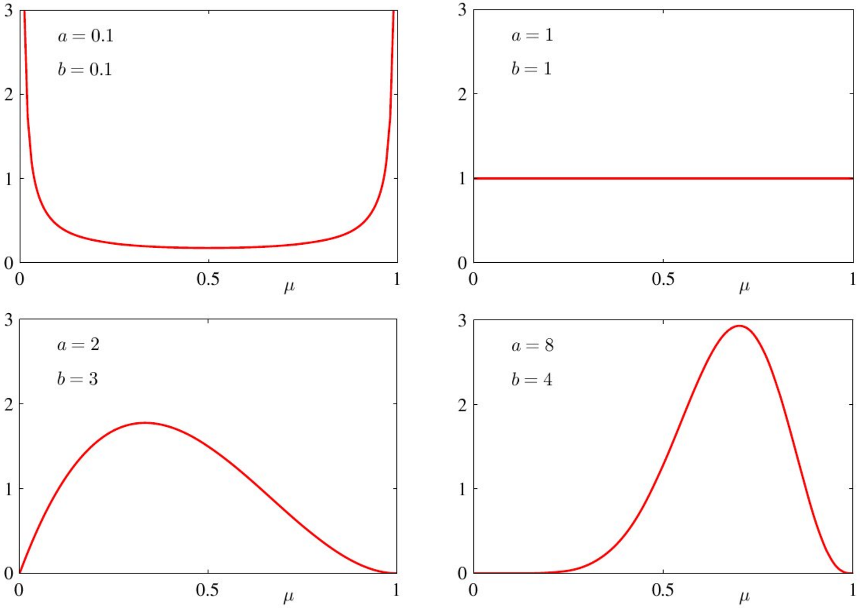
Or $\text{Beta}(\theta, 1, 1) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} =$ constante

Max de vraisemblance $\Longleftrightarrow$ Max a posteriori avec a priori uniforme

$n \to +\infty \Longrightarrow$ max de vraisemblance $\approx$ max a posteriori

$\Longrightarrow$ l'a priori devient négligeable

La loi Beta



Loi normale et loi conjuguée

- fonction de vraisemblance = loi normale, σ^2 connue

  $\Longrightarrow$ loi a priori conjuguée : loi Γ

La loi Γ

- $X \sim \Gamma(x, k, \theta)$
- fonction de densité de la loi Γ :

$$f(x, k, \theta) = x^{k-1} \frac{e^{-x/\theta}}{\theta^k \Gamma(k)} \quad \forall x, k, \theta > 0$$

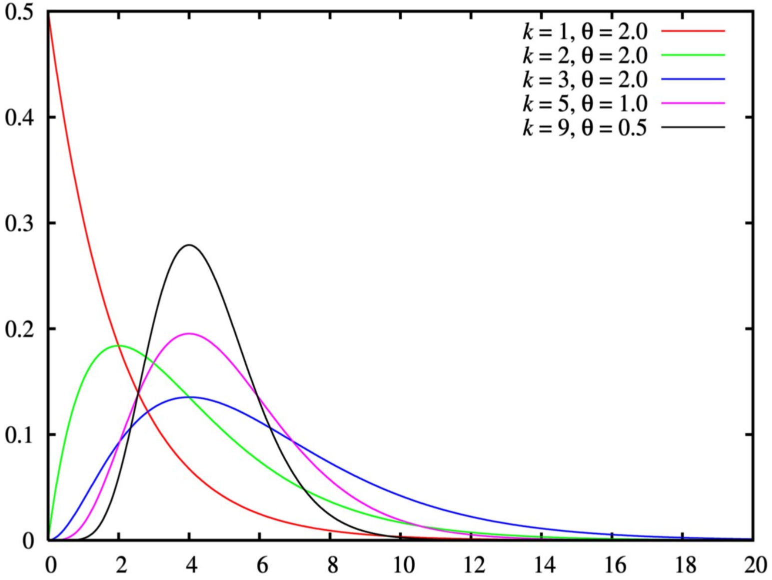
- $\Gamma(k) = \int_0^{+\infty} t^{k-1} e^{-t} dt$
- $E(X) = k\theta, \quad V(X) = k\theta^2$

Lorsque k entier : $\Gamma(x, k, \theta) =$ loi de k variables indépendantes suivant une loi exponentielle d'espérance θ

- Familles de lois conjuguées :

  `http ://en.wikipedia.org/wiki/Conjugate_prior`

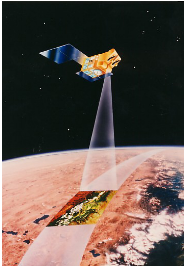
Loi Gamma



k = 1, θ = 2.0
k = 2, θ = 2.0
k = 3, θ = 2.0
k = 5, θ = 1.0
k = 9, θ = 0.5
0.5
0.4
0.3
0.2
0.1
0
0
2
4
6
8
10
12
14
16
18
20

- 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 (2/3)

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

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

- nouvelle image envoyée par SPOT5 :

- zone Z : niveau de gris $= n = 80$
- Connaissance a priori : 60% de PI, 10% de PPI, 30% de NI

Problème : zone $Z = PI$ ou PPI ?

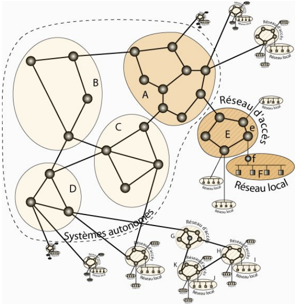
Prévention des risques d'inondation (3/3)

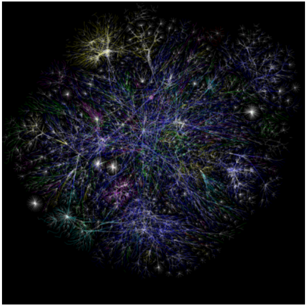
Problème : zone $Z = PI$ ou PPI ?

- 2 hypothèses :
  1. $\theta_1 =$ « Z est de type PI »
  2. $\theta_2 =$ « Z est de type PPI »
- Idée : calcul du MAP d'obtenir la zone Z sous θ_1 ou sous θ_2
- $\pi(\theta_1|\mathbf{x}) = \dfrac{L(\mathbf{x},\theta_1)\pi(\theta_1)}{\sum_{\theta\in\Theta} L(\mathbf{x},\theta)\pi(\theta)}$ $\qquad$ $\pi(\theta_2|\mathbf{x}) = \dfrac{L(\mathbf{x},\theta_2)\pi(\theta_2)}{\sum_{\theta\in\Theta} L(\mathbf{x},\theta)\pi(\theta)}$
- Rappel cours 4 : $L(\mathbf{x},\theta_1) \approx 0,0121$ $\qquad$ $L(\mathbf{x},\theta_2) \approx 0,0484$
- a priori : $\pi(\theta_1) = 0,6$ $\qquad$ $\pi(\theta_2) = 0,1$
- $\pi(\theta_1|\mathbf{x}) = \dfrac{0,0121 \times 0,6}{\sum_{\theta\in\Theta} L(\mathbf{x},\theta)\pi(\theta)}$ $\qquad$ $\pi(\theta_2|\mathbf{x}) = \dfrac{0,0484 \times 0,1}{\sum_{\theta\in\Theta} L(\mathbf{x},\theta)\pi(\theta)}$

MAP $\Longrightarrow$ parcelle inondable (PI)

- Réseau informatique : transfert de paquets





- Problème : analyse des paquets perdus sur un sous-réseau
- X : variable aléatoire « nombre de paquets envoyés jusqu’à bonne réception »
- X loi géométrique : $P(X = n) = (1 - p)^{n-1} p$

  p : probabilité qu’un paquet soit correctement transmis

- observation de 7 réalisations de X :

2	3	8	3	4	7	8

Estimation de p ?

1. estimation par max de vraisemblance
2. estimation par MAP

Analyse d'un trafic réseau (3/4)

1 estimation par max de vraisemblance

2	3	8	3	4	7	8

- vraisemblance : $L(\mathbf{x}, \theta) = \prod_{i=1}^{7} P(x_i|\theta)$

$$\theta = \text{estimation de } p$$

- observations $\Longrightarrow L(\mathbf{x}, \theta) = (1-\theta)^{28}\theta^7$

$$\Longrightarrow \ln L(\mathbf{x}, \theta) = 28 \ln(1-\theta) + 7 \ln \theta$$

$$\Longrightarrow \frac{\partial \ln L(\mathbf{x}, \theta)}{\partial \theta} = \frac{-28}{1-\theta} + \frac{7}{\theta} = \frac{7 - 35\theta}{p(1-\theta)}$$

$$\Longrightarrow \text{maximum de vraisemblance} = \theta = 0,2$$

2 estimation par max de vraisemblance

2	3	8	3	4	7	8

- A priori : $\pi(\theta) = \text{Beta}(\theta, 2, 15) = \dfrac{\Gamma(17)}{\Gamma(2)\Gamma(15)}\theta^1(1-\theta)^{14}$
- $\text{Argmax}_\theta \pi(\theta|\mathbf{x}) = \text{Argmax}_\theta L(\mathbf{x}, \theta)\pi(\theta)$

$$= \text{Argmax}_\theta [(1-\theta)^{28}\theta^7] \times [(1-\theta)^{14}\theta]$$

$$= \text{Argmax}_\theta (1-\theta)^{42}\theta^8$$

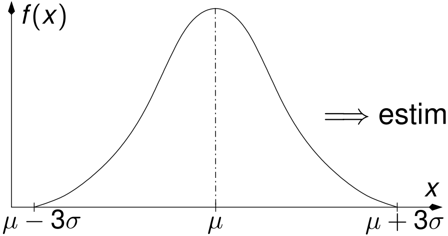
$$= \text{Argmax}_\theta\ 42\ln(1-\theta) + 8\ln\theta$$

$\Longrightarrow \theta_{MAP} = 0,16$

2 Estimation de densité

En route vers le non paramétrique

- jusqu'à maintenant : estimations des paramètres des densités :
  - max de vraisemblance
  - max a posteriori



$\Longrightarrow$ estimation de μ, σ^2

$\Longrightarrow$ *statistique paramétrique*

$\Longrightarrow$ connaissance de la famille de lois a priori

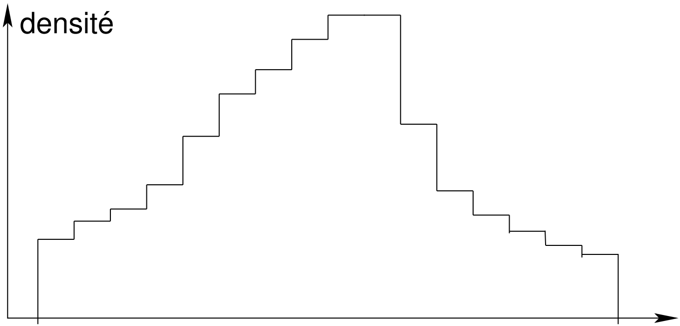
Que faire si aucune loi a priori évidente ?

$\Longrightarrow$ statistique non paramétrique

Estimation de densité : idée naïve

- X : variable aléatoire
- $\{x_1, \dots, x_n\}$: les classes de X
- échantillon de taille $n \Longrightarrow n_i$ individus dans la classe x_i

Densité $\approx$ histogramme des valeurs observées

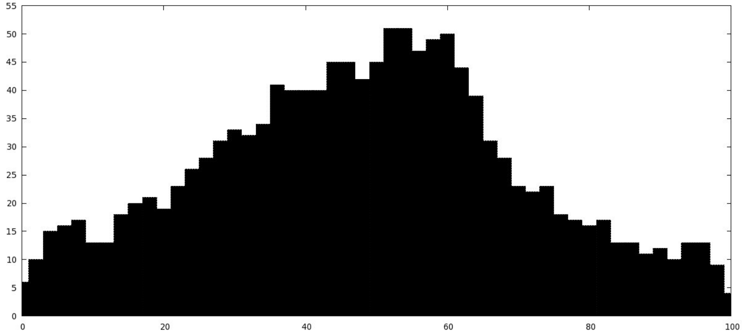


Problème : fonction de densité très irrégulière

Estimation de densité : la fenêtre mobile (1/2)

Idée de la fenêtre mobile

- X : variable aléatoire
- $\forall x$ valeur de X, $I_x = [x - \frac{h}{2}, x + \frac{h}{2}[$ = classe de longueur h
- densité estimée : $\hat{f}(x) = \frac{1}{n} \# x_i \ : x - \frac{h}{2} \leq x_i < x + \frac{h}{2}$



Problème : fonction de densité encore trop irrégulière densité estimée : $\hat{f}(x) = \frac{1}{n} \# x_i \ : x - \frac{h}{2} \leq x_i < x + \frac{h}{2}$

$$\Longrightarrow \hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

où $K =$ fonction indicatrice de l'intervalle $\left[-\frac{1}{2}, \frac{1}{2}\right[$:

$$K(u) = \begin{cases} 0 & \text{si } u \geq \frac{1}{2} \text{ ou } u < -\frac{1}{2} \\ 1 & \text{si } -\frac{1}{2} \leq u < \frac{1}{2} \end{cases}$$

Estimation de densité par la méthode du noyau

- X : variable aléatoire
- $\forall x$ valeur de X, $I_x = [x - \frac{h}{2}, x + \frac{h}{2}[$ = classe de longueur h
- densité estimée : $\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$

  avec $K =$ fonction continue = *noyau*

Estimation de densité par la méthode du noyau

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

- noyaux fréquemment utilisés :
  - noyau gaussien : $K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$
  - noyau d'Epanechnikov : $K(u) = \frac{3}{4\sqrt{5}}\left(1 - \frac{u^2}{5}\right)$ pour $|u| < \sqrt{5}$
- constante $h =$ constante de lissage

$$\begin{cases} h \text{ petit} & \Longrightarrow \hat{f} \text{ très irrégulière} \\ h \text{ grand} & \Longrightarrow \hat{f} \text{ très (trop) lisse} \end{cases}$$

- h souvent défini empiriquement