MAPSI — cours 3 : Maximum de vraisemblance Maximum a posteriori

Pierre-Henri Wuillemin & Christophe Gonzales

LIP6 / ISIR – Sorbonne Université, France

Plan du cours n°3

1. Vraisemblance et prise de décision
2. Estimation par maximum de vraisemblance (ML)
3. Estimation par maximum a posteriori (MAP)

2. Vraisemblance et prise de décision

Vraisemblance d'un échantillon : loi discrète connue

- Échantillon $\mathbf{x} = (x_1, \ldots, x_n)$ 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

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

l'hypothèse i.i.d est essentielle !

Vraisemblance d'un échantillon dans le cas discret

- $L(\mathbf{x})$ = Vraisemblance de l'échantillon
- $L(\mathbf{x})$ = proba d'obtenir cet échantillon sachant la loi P

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

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

$\Longrightarrow$ hypothèse i.i.d. vérifiée

- échantillon 1 :

P	P	F	F	P	P	F	P	P	P

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

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

- échantillon 2 :

F	F	P	P	F	F	P	F	F	F

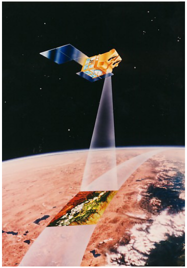
$$\Longrightarrow L(\mathbf{x}) = \prod_{i=1}^{3} P(\text{Pile}) \times \prod_{i=1}^{7} P(\text{Face})$$

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

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 ?

Prévention des risques d'inondation (3/4)

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 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

2 ML : Estimation par maximum de vraisemblance

- Paramètre à estimer : Θ

Exemple 1 :   $X \in \{\text{pile,face}\}$

$$P(X) = \begin{array}{cc} \text{pile} & \text{face} \\ \hline \multicolumn{1}{|c|}{\theta_1} & \multicolumn{1}{c|}{\theta_2} \\ \hline \end{array} \Longrightarrow \Theta = \{\theta_1, \theta_2\}$$

Exemple 2 : recommandations : $r_A \in \{1, 2, 3\}, r_B \in \{a, b\}$

$$P(r_A, r_B) = \begin{array}{c|c|c|} & a & b \\ \hline 1 & \theta_1 & \theta_2 \\ \hline 2 & \theta_3 & \theta_4 \\ \hline 3 & \theta_5 & \theta_6 \\ \hline \end{array} \Longrightarrow \Theta = \{\theta_1, \dots, \theta_6\}$$

Apprentissage par vraisemblance : le cas discret

- Paramètre à estimer : Θ
- Échantillon $\mathbf{x} = (x_1, \ldots, x_n)$ 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

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

Vraisemblance d'un échantillon dans le cas discret

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

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

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

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

Vraisemblance d'un échantillon dans le cas continu

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

Apprentissage de Θ par vraisemblance

- pièce de monnaie : $P(\text{Pile}) = \theta_1 = ???$ et $P(\text{Face}) = \theta_2 = ???$
- paramètre Θ = proba de Pile $= \theta_1 = ???$
- échantillon :

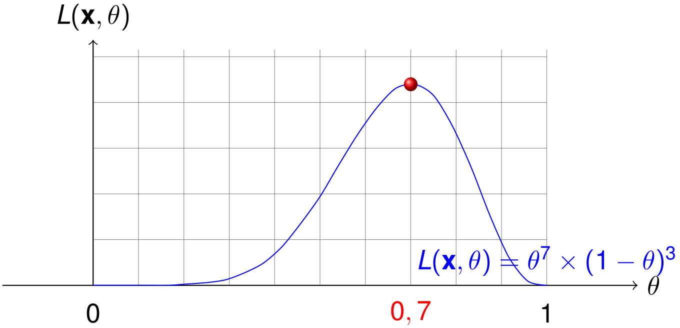
P	P	F	F	P	P	F	P	P	P

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

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

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

Apprentissage de Θ par vraisemblance



solution optimale : $\theta = 0,7$

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 ?

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

- 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 $\mathbf{x} = \{x_1, \ldots, x_n\}$
- $P(\mathbf{x}|p) = \prod_{i=1}^{n} P(x_i|p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$

Problème : à partir de $\mathbf{x}$, 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 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)

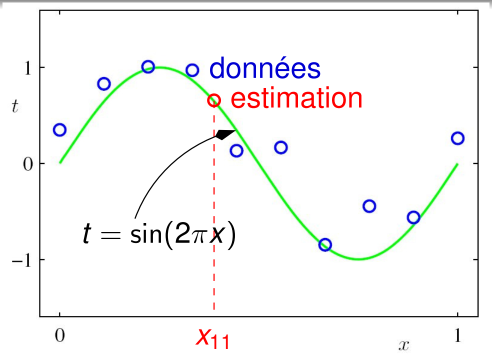
- $X \sim \mathcal{N}(\mu, \sigma^2)$
- paramètre $\Theta = (\mu, \sigma^2)$
- Log vraisemblance :

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

- 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



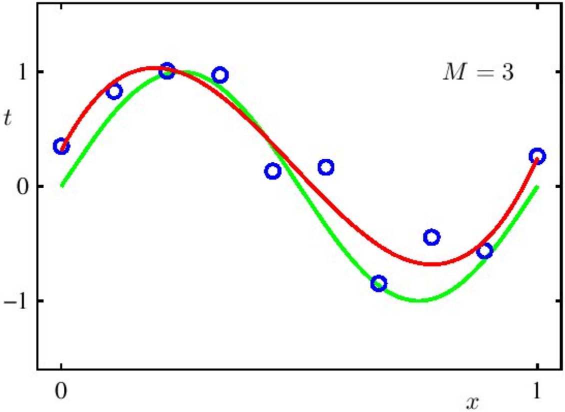
Observations
(x_1, t_1) $\vdots$ (x_{10}, t_{10})

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

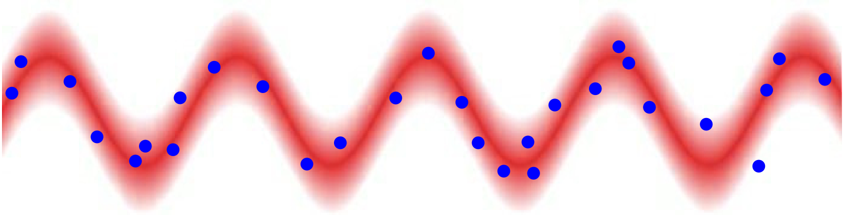
$\Longrightarrow$ reconnaissance de la courbe verte

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

Problème d'ajustement (4/6)

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

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

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

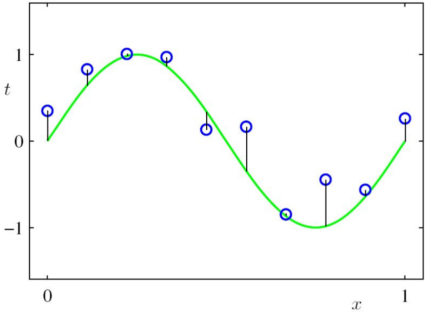
$$= \prod_{i=1}^{n} \mathcal{N}(t_i|y(x_i, \mathbf{w}), \sigma^2)$$

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



$\Longrightarrow$ Moindres carrés

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

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

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

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

2. MAP : Estimation par maximum a posteriori

Retour sur la loi binomiale

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

- 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

Maximum a posteriori $\Longrightarrow$ modèle bayésien

Modèle bayésien

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

- $\mathcal{X} =$ l'espace des observations (échantillons) $\mathbf{x}$ 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 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 :

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

Le modèle bayésien (4/4)

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

- 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 $\Longrightarrow \pi(\mathbf{x}|\theta) = L(\mathbf{x}, \theta) = \begin{cases} \prod_{i=1}^n P(x_i|\theta) & \text{(discret)} \\ \prod_{i=1}^n p(x_i|\theta) & \text{(continu)} \end{cases}$

MAP : retour sur la pièce de monnaie (1/6)

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

  $0 \Longleftrightarrow$ Face  $\quad 1 \Longleftrightarrow$ 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 : $\Theta = \{\theta_1 = 1, \theta_2 = 2/3, \theta_3 = 1/2, \theta_4 = 1/3\}$
- Info a priori : $\pi(\theta_1) = \frac{1}{32}, \pi(\theta_2) = \frac{1}{4}, \pi(\theta_3) = \frac{1}{2}, \pi(\theta_4) = \frac{7}{32}$

Problème : quelle est la valeur du maximum a posteriori ?

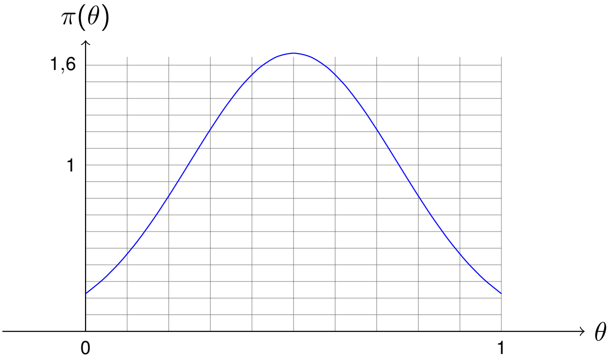
- Modèle bayésien : $\Theta = \{\theta_1 = 1, \theta_2 = 2/3, \theta_3 = 1/2, \theta_4 = 1/3\}$
- $L(\mathbf{x}, \theta_1) = \pi(\mathbf{x}|\theta_1) = \prod_{i=1}^{3} P(x_i|\theta_1) = 1^3 \times 0^0 = 1$
- $L(\mathbf{x}, \theta_2) = \pi(\mathbf{x}|\theta_2) = \prod_{i=1}^{3} P(x_i|\theta_2) = \frac{2}{3}^3 \times \left(1 - \frac{2}{3}\right)^0 = \frac{2}{3}^3 \approx 0,296$
- $L(\mathbf{x}, \theta_3) = \pi(\mathbf{x}|\theta_3) = \prod_{i=1}^{3} P(x_i|\theta_3) = \frac{1}{2}^3 \times \left(1 - \frac{1}{2}\right)^0 = \frac{1}{2}^3 0,125$
- $L(\mathbf{x}, \theta_4) = \pi(\mathbf{x}|\theta_4) = \prod_{i=1}^{3} P(x_i|\theta_4) = \frac{1}{3}^3 \times \left(1 - \frac{1}{3}\right)^0 = \frac{1}{3}^3 \approx 0,037$

- Info a priori : $\pi(\theta_1) = \frac{1}{32}$, $\pi(\theta_2) = \frac{1}{4}$, $\pi(\theta_3) = \frac{1}{2}$, $\pi(\theta_4) = \frac{7}{32}$
- $\pi(\theta_1|\mathbf{x}) = \dfrac{L(\mathbf{x},\theta_1)\pi(\theta_1)}{\sum_{\theta\in\Theta} L(\mathbf{x},\theta)\pi(\theta)} \propto 1 \times \dfrac{1}{32} = 0,03125$
- $\pi(\theta_2|\mathbf{x}) = \dfrac{L(\mathbf{x},\theta_2)\pi(\theta_2)}{\sum_{\theta\in\Theta} L(\mathbf{x},\theta)\pi(\theta)} \propto \dfrac{2}{3}^{3} \times \dfrac{1}{4} \approx 0,074$
- $\pi(\theta_3|\mathbf{x}) = \dfrac{L(\mathbf{x},\theta_3)\pi(\theta_3)}{\sum_{\theta\in\Theta} L(\mathbf{x},\theta)\pi(\theta)} \propto \dfrac{1}{2}^{3} \times \dfrac{1}{2} = 0,0625$
- $\pi(\theta_4|\mathbf{x}) = \dfrac{L(\mathbf{x},\theta_4)\pi(\theta_4)}{\sum_{\theta\in\Theta} L(\mathbf{x},\theta)\pi(\theta)} \propto \dfrac{1}{3}^{3} \times \dfrac{7}{32} \approx 0,008$

Max a posteriori : $\Theta = \theta_2 \Longrightarrow X \sim \mathcal{B}(1,\theta_2) = \mathcal{B}(1,2/3)$

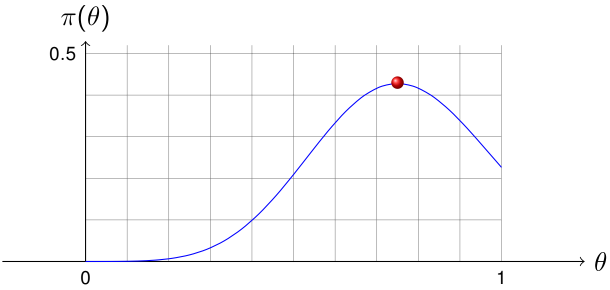
⚠ probabilité que la pièce tombe sur Face $\neq 0$

- Modèle bayésien : $\Theta \in [0, 1]$
- Info a priori : $\Theta \sim$ loi normale tronquée ($\mu = 1/2, \sigma = 1/4$) :
- densité : $\pi(\theta) = \begin{cases} \frac{1}{0{,}9544} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{\theta-\mu}{\sigma}\right)^2\right) & \text{si } \theta \in [0,1] \\ 0 & \text{sinon} \end{cases}$



$$\pi(\theta|\mathbf{x}) = \frac{L(\mathbf{x},\theta)\pi(\theta)}{\int_{\theta'\in\Theta} L(\mathbf{x},\theta')\pi(\theta')} \propto L(\mathbf{x},\theta)\pi(\theta) = \theta^3 \times \pi(\theta)$$

$$\propto \begin{cases} \theta^3 \times \frac{1}{0{,}9544}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2}\left(\frac{\theta-\mu}{\sigma}\right)^2\right) & \text{si } \theta \in [0,1] \\ 0 & \text{sinon} \end{cases}$$



solution optimale : $\theta = 0,75$

$$\pi(\theta|\mathbf{x}) \propto \theta^3 \times \frac{1}{0{,}9544}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2}\left(\frac{\theta-\mu}{\sigma}\right)^2\right) \text{ pour } \theta \in [0,1]$$

$$\Longrightarrow \log\pi(\theta|\mathbf{x}) = 3\log\theta - \frac{1}{2}\left(\frac{\theta-\mu}{\sigma}\right)^2 + \text{ constante}$$

$$\Longrightarrow \frac{\partial \log\pi(\theta|\mathbf{x})}{\partial\theta} = \frac{3}{\theta} - \frac{\theta-\mu}{\sigma^2}$$

$$\Longrightarrow \frac{\partial \log\pi(\theta|\mathbf{x})}{\partial\theta} = 0 \Leftrightarrow \theta^2 - \mu\theta - 3\sigma^2 = 0$$

$$\Longrightarrow \theta = 0,75$$

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

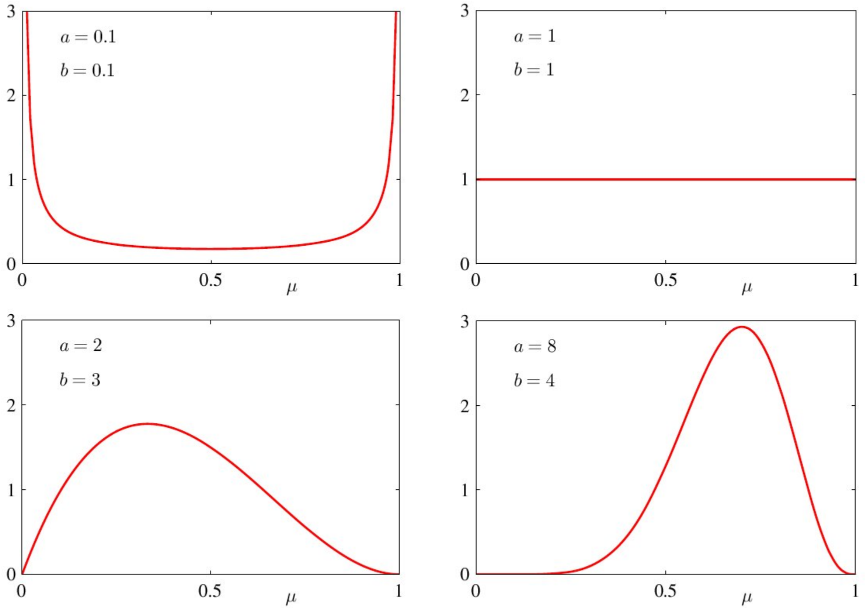
$\Longrightarrow$ loi Beta et loi binomiales conjuguées

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 : $\pi(\theta|\mathbf{x}) \propto \theta^{x+a-1}(1-\theta)^{b+n-x-1}$

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

La loi Beta



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

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 :

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

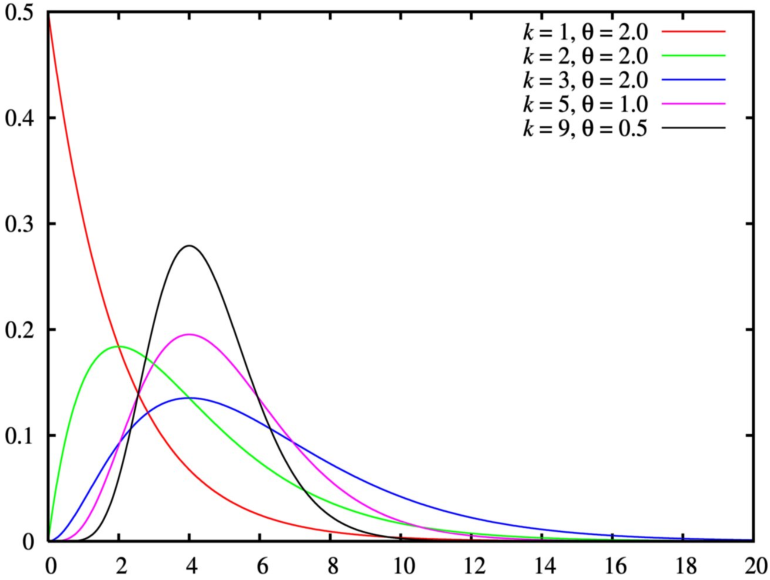
- $\Gamma(k) = \int_0^{+\infty} t^{k-1} e^{-t} dt$
- $E(X) = k\theta, \quad V(X) = k\theta^2$

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

- Familles de lois conjuguées :

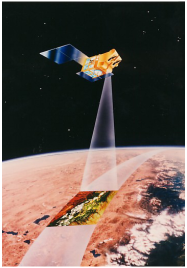
  http ://en.wikipedia.org/wiki/Conjugate_prior

Loi Gamma


$k = 1, \theta = 2.0$
$k = 2, \theta = 2.0$
$k = 3, \theta = 2.0$
$k = 5, \theta = 1.0$
$k = 9, \theta = 0.5$
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*)

- 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

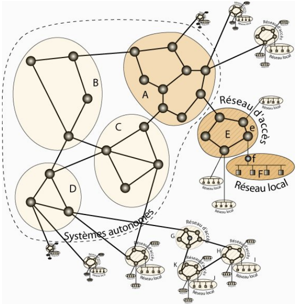
Problème : zone $Z = PI$ ou PPI ?

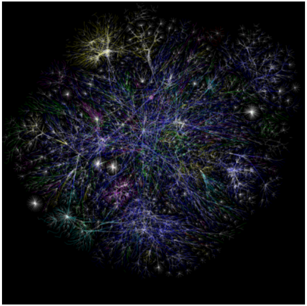
Problème : zone $Z = PI$ ou PPI ?

- 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

Analyse d'un trafic réseau (2/4)

- observation de 7 réalisations de X :

2	3	8	3	4	7	8

Estimation de p ?

1. estimation par ML
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 : $L(\mathbf{x}, \theta) = \prod_{i=1}^{7} P(x_i|\theta)$

  $\theta =$ estimation de p

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

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

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

  $\Longrightarrow$ maximum de vraisemblance $= \theta = 0,2$

Analyse d'un trafic réseau (4/4)

2 estimation par max de vraisemblance

2	3	8	3	4	7	8

- A priori : $\pi(\theta) = \text{Beta}(\theta, 2, 15) = \dfrac{\Gamma(17)}{\Gamma(2)\Gamma(15)}\theta^1(1-\theta)^{14}$
- $$\begin{aligned}\text{Argmax}_\theta \pi(\theta|\mathbf{x}) &= \text{Argmax}_\theta L(\mathbf{x}, \theta)\pi(\theta) \\ &= \text{Argmax}_\theta [(1-\theta)^{28}\theta^7] \times [(1-\theta)^{14}\theta] \\ &= \text{Argmax}_\theta (1-\theta)^{42}\theta^8 \\ &= \text{Argmax}_\theta\ 42\ln(1-\theta) + 8\ln\theta\end{aligned}$$

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