Chapitre 3

Inférence bayésienne : l'approche théorie de la décision

3.1 Théorie de la décision et fonction coût

Un modèle statistique met en jeu trois espaces : espace des *observations* $\mathcal{X}$, espace des *paramètres* Θ, et espace des *décisions* $\mathcal{D}$. Pour une décision donnée d concernant un paramètre θ on définit une fonction coût :

Définition 5 On appelle coût tout fonction $C(\theta, d)$ de $\Theta \times \mathcal{D}$ dans $\mathbb{R}^+$.

L'inférence décisionelle consiste à prendre une décision $d \in \mathcal{D}$ concernant $\theta \in \Theta$ au vu d'une observation $\boldsymbol{x} \in \mathcal{X}$, $\boldsymbol{x}$ et θ étant reliés par la loi $p(\boldsymbol{x}|\theta)$. La théorie de la décision ajoute l'hypothèse supplémentaire que les actions (ou les décisions) peuvent être évaluées. Chaque action (ou décisions) conduit donc à une récompse r, d'utilité $U(r)$. On définit alors une *fonction d'utilité*

$$U(\theta, d) = \mathrm{E}_{\theta,d}\left\{U(r)\right\}$$

et on en déduit une *fonction de coût*

$$C(\theta, d) = -U(\theta, d).$$

On peut alors définir une *fonction risque moyen* (risque fréquentiste)

$$\begin{aligned} R(\theta, \delta) &= \mathrm{E}_{x|\theta}\left\{C(\theta, \delta(\boldsymbol{x}))\right\} \\ &= \int_{\mathcal{X}} C(\theta, \delta(\boldsymbol{x})) p(\boldsymbol{x}|\theta)\,\mathrm{d}\boldsymbol{x}, \end{aligned}$$

une *fonction coût moyen a posteriori*

$$\begin{aligned} \bar{C}(d|\boldsymbol{x}) &= \mathrm{E}_{\theta|x}\left\{C(\theta, d)\right\} \\ &= \int_{\Theta} C(\theta, d) p(\theta|\boldsymbol{x})\,\mathrm{d}\theta, \end{aligned}$$

et finalement, le *risque de Bayes*,

$$\begin{aligned} r(\delta) &= \mathrm{E}_{\theta,x}\left\{C(\theta, \delta(\boldsymbol{x}))\right\} \\ &= \int_{\Theta}\int_{\mathcal{X}} C(\theta, \delta(\boldsymbol{x})) p(\boldsymbol{x}|\theta) p(\theta)\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\theta \end{aligned}$$

une valeure réelle et non plus une fonction de θ qui permet la comparaison direct des estimateurs.
Dans l'approche classique, le choix d'un estimateur n'est pas trés simple. On définit un estimateur $\delta(\boldsymbol{x})$, fonction des observations $\boldsymbol{x}$, on en déduite l'espace de tous les estimateurs possibles $\mathcal{D}$ et la loi $p(\delta(\boldsymbol{x})|\theta)$ à partir de l'espace des observations $\mathcal{X}$et la loi des observations $p(\boldsymbol{x}|\theta)$, et, finalement on définit le risque moyen

$$R(\theta,\delta) = \int_{\mathcal{D}} C(\theta,\delta(x)) p(\delta|\theta)\,\mathrm{d}\delta.$$

et le *risque minimax*:

$$R_{\text{MinMax}} = \inf_{\delta\in\mathcal{D}} \sup_{\theta\in\Theta} R(\theta,\delta)$$

On choisit ensuite, un *estimateur minimax* $\delta^*(\boldsymbol{x})$ tel que

$$\sup_{\theta\in\Theta} R(\theta,\delta^*) = R_{\text{MinMax}}.$$

Heureusement, dans l'approche bayésienne la procédure est plus simple : On choisit comme estimateur $\delta(\boldsymbol{x})$ la fonction qui minimise le coût moyen

$$\delta(\boldsymbol{x}) = \arg\min_{\mathcal{D}} \left\{\bar{C}(d|\boldsymbol{x})\right\}.$$

Le paragraphe qui suit précisera ce point.

3.2 Fonction coût et estimateurs bayésiens

Une inférence bayésienne est ainsi fondée sur la détermination rigoureuse de trois facteurs :

1. la loi des observations $p(\boldsymbol{x}|\theta)$
2. la distribution *a priori* des paramètres $p(\theta)$
3. le coût associé aux décisions $C(\theta,\delta(\boldsymbol{x}))$

La détermination de ces trois facteurs se pose de difficultés comparables et ne peuvent se faire que par des considération partiellement subjectives. Notons que les critiques fréquents de l'approche bayésienne se font souvent sur le point 2, alors que, conceptuellement les points 1 et 3 se trouvent sur le même rang et que le point 3 qui est un point commun entre l'approche classique et l'approche bayésienne est encore plus subtile.

Définition 6 [Estimateur de Bayes] On appelle *estimateur de Bayes* associé à une loi *a posteriori* $p(\theta|\boldsymbol{x})$ et à une fonction coût $C(\theta,\delta(\boldsymbol{x}))$, tout estimateur δ qui minimise le risque de Bayes $r(\delta)$. Cet estimateur peut être obtenu en choisissant, pour chaque $\boldsymbol{x}\in\mathcal{X}$, la valeur $\delta(\boldsymbol{x})$ qui minimise le coût a posteriori, $\bar{C}(d|\boldsymbol{x})$.

La difficultée réside dans le choix ou la détermination de la fonction coût, qui n'est pas propre de l'approche bayésienne. Bien qu'il existe des arguments pour ce choix, nous n'allons pas le développer ici. Je propose plutôt de faire l'inventaire d'un certain nombre de fonctions coûts usuelles.

3.3 Fonctions de coûts usuelles et les estimateurs associés

Le choix d'une fonction coût dépend de l'application concernée. Lorque le contexte ne permet pas la détermination de la fonction d'utilité qui permet d'en déduire la fonction coût, on peut avoir recours à des fonctions coûts classiques, qui sont à la fois simples et bien étudiées.

3.3.1 Coût quadratique

Parmi les fonctions coûts la plus utilisée on peut noter le coût quadratique

$$C(\theta, d) = (\theta - d)^2.$$

qui pénalise trop fortement les grandes erreurs, du fait de sa convexité stricte. Parmis les propriétés de cette fonction coût et l'estimateur bayésien qui en découlent on peut citer les suivantes :

Proposition 1 L'estimateur de Bayes δ de $\theta \in \mathbb{R}$, associé à la loi *a posteriori* $p(\theta|\boldsymbol{x})$ et au coût quadratique $C(\theta, d) = (\theta - d)^2$, est la moyenne a posteriori

$$\delta(\boldsymbol{x}) = \mathrm{E}\left[\theta|\boldsymbol{x}\right] = \int \theta p(\theta|\boldsymbol{x})\, \mathrm{d}\theta.$$

En effet on a

$$\mathrm{E}\left[C(\theta, d)\right] = d^2 - 2d\mathrm{E}\left[\theta\right] + \mathrm{E}\left[\theta^2\right] = (d - \mathrm{E}\left[\theta\right])^2 + \mathrm{Var}\left[\theta\right],$$

qui est minimisé lorsque $d = \mathrm{E}\left[\theta\right]$.

Corollaire 1.1 L'estimateur de Bayes $\delta(\boldsymbol{x})$, associé à $p(\theta|\boldsymbol{x})$ et au coût quadratique pondéré

$$C(\theta, d) = \omega(\theta)\,(\theta - d)^2,$$

est

$$\delta(\boldsymbol{x}) = \frac{\mathrm{E}\left[\omega(\theta)\,\theta|\boldsymbol{x}\right]}{\mathrm{E}\left[\omega(\theta)|\boldsymbol{x}\right]} = \frac{\int \omega(\theta)\theta p(\theta|\boldsymbol{x})\, \mathrm{d}\theta}{\int \omega(\theta) p(\theta|\boldsymbol{x})\, \mathrm{d}\theta}.$$

Corollaire 1.2 L'estimateur de Bayes δ, associé à $p(\theta|\boldsymbol{x})$ et au coût quadratique généralisé

$$C(\theta, d) = q(\theta - d)^2$$

est la moyenne a posteriori

$$\delta(\boldsymbol{x}) = \mathrm{E}\left[\theta|\boldsymbol{x}\right] = \int \theta p(\theta|\boldsymbol{x})\, \mathrm{d}\theta.$$

Notons que l'estimateur associé à cette fonction ne dépend pas du coefficient q.
Ce résultat s'étend facilement au cas vectoriel $\boldsymbol{\theta}$:

Corollaire 1.3 L'estimateur de Bayes $\boldsymbol{\delta}$, associé à $p(\boldsymbol{\theta}|\boldsymbol{x})$ et au coût quadratique généralisé

$$C(\boldsymbol{\theta}, \boldsymbol{d}) = (\boldsymbol{\theta} - \boldsymbol{d})^t \boldsymbol{Q} (\boldsymbol{\theta} - \boldsymbol{d})$$

est la moyenne a posteriori

$$\boldsymbol{\delta}(\boldsymbol{x}) = \mathrm{E}\left[\boldsymbol{\theta}|\boldsymbol{x}\right] = \int \boldsymbol{\theta} p(\boldsymbol{\theta}|\boldsymbol{x})\, \mathrm{d}\boldsymbol{\theta}.$$

Cet estimateur ne dépend pas de la matrice $\boldsymbol{Q}$, sous réserve que cette matrice soit définie positive.

3.3.2 Coût absolu

Une autre fonction coût trés utilisée est

$$C(\theta, d) = |\theta - d|$$

ou, plus généralement,

$$C(\theta, d) = \begin{cases} k_2(\theta - d) & \text{si } \theta \leq d, \\ k_1(d - \theta) & \text{si } \theta \geq d. \end{cases}$$

Ces fonctions restent convexes, mais augmentent moins rapidement que les coûts quadratiques. Ceci permet d'éviter de trop pénaliser les grandes erreurs. L'usage de ces fonctions pose quelques difficultées mathématiques lors de leur optimisations. Ceci est essentiellement du à la discontinuité de la dérivée de ces fonctions à l'origine. C'est pourquoi, d'autres fonctions qui mélangent les coûts absolu et quadratique sont parfois utilisées, comme par exemple

$$C(\theta, d) = \begin{cases} (\theta - d)^2 & \text{si } |\theta - d| < t, \\ 2t|\theta - d| - t^2 & \text{sinon .} \end{cases}$$

Proposition 2 L'estimateur de Bayes $\delta(\boldsymbol{x})$ de $\theta \in \mathbb{R}$, associé à la loi $p(\theta|\boldsymbol{x})$ et au coût absolu $C(\theta, d) = |\theta - d|$, est la médiane de la loi a posteriori $p(\theta|\boldsymbol{x})$

$$\delta(\boldsymbol{x}) = a | P(\theta < a|\boldsymbol{x}) = P(\theta > a|\boldsymbol{x}) = \frac{1}{2}.$$

Proposition 3 L'estimateur de Bayes δ de $\theta \in \mathbb{R}$, associé à la loi $p(\theta|\boldsymbol{x})$ et au coût absolu généralisé

$$C(\theta, d) = \begin{cases} k_2(\theta - d) & \text{si } \theta < a, \\ k_1(d - \theta) & \text{sinon} \end{cases},$$

est un fractile d'ordre $(k_2/(k_1 + k_2))$ de la loi *a posteriori* $p(\theta|\boldsymbol{x})$

$$\delta(\boldsymbol{x}) = a | k_1 P(\theta < a|\boldsymbol{x}) = k_2 P(\theta > a|\boldsymbol{x})$$

ou

$$\delta(\boldsymbol{x}) = a | P(\theta < a|\boldsymbol{x}) = \frac{k_2}{k_1 + k_2}.$$

3.3.3 Coût "0-1"

Lorsque la décision est binaire, comme par exemple en tests d'hypothèse $\mathcal{D} = \{0, 1\}$, où $d = 1$ correspond à l'acceptation de l'hypothèse $H_0 : \theta \in \Theta_0$ et $d = 1$ correspond à l'acceptation de l'hypothèse $H_1 : \theta \notin \Theta_0$, une fonction coût qui permet d'estimer le nombre de faux alarms est la fonction coût

$$C(\theta, d) = \begin{cases} 1 - d & \text{si } \theta \in \Theta_0, \\ d & \text{sinon.} \end{cases}$$

En effet, si on calcul le coût moyen

$$\begin{aligned} \bar{C}(\delta) &= \mathrm{E}_{\theta|x}\left\{C(\theta, \delta(\boldsymbol{x}))\right\} \\ &= \int_\Theta C(\theta, \delta(\boldsymbol{x})) p(\theta|\boldsymbol{x}) \, \mathrm{d}\theta \\ &= \begin{cases} \Pr\left\{\delta(\boldsymbol{x}) = 0\right\} & \text{si } \theta \in \Theta_0, \\ \Pr\left\{\delta(\boldsymbol{x}) = 1\right\} & \text{sinon.} \end{cases} \end{aligned}$$

ce qui correspondent aux *erreurs de première et de seconde espèces* utilisées en statistque classique.

Proposition 4 L'estimateur de Bayes, associé à la loi $p(\theta|\boldsymbol{x})$ et au coût "0-1" est

$$\delta(\boldsymbol{x}) = \begin{cases} 1 & \text{si } P(\theta \in \Theta|\boldsymbol{x}) > P(\theta \notin \Theta|\boldsymbol{x}), \\ 0 & \text{sinon.} \end{cases}$$

Lors de l'estimation on définit la fonction coût

$$C(\theta, d) = \begin{cases} 0 & \text{si } |d - \theta| < b, \\ 1 & \text{sinon} \end{cases},$$

dont sa moyenne est

$$\mathrm{E}[C] = \Pr\{|d - \theta| > b\}.$$

L'estimateur optimal correspondant est le centre de l'interval $[-b, b]$ qui a la probabilité maximale.

3.3.4 Coût "Dirac"

Dans les cas de l'estimation bayésien aussi on peut définir

$$C(\theta, d) = 1 - u(\theta - d),$$

où $u(t)$ est la distribution de Dirac.

Proposition 5 L'estimateur de Bayes, associé à la loi *a posteriori* $p(\theta|\boldsymbol{x})$ et au coût "Dirac", définit précédament, est

$$\delta(\boldsymbol{x}) = \arg\max_{\theta} \{p(\theta|\boldsymbol{x})\}.$$

En effet, on a

$$\begin{aligned} \delta(\boldsymbol{x}) &= \arg\min_{d} \left\{ \int C(\theta, d)\, p(\theta|\boldsymbol{x})\, \mathrm{d}\theta \right\} \\ &= \arg\min_{d} \left\{ \int [1 - u(\theta - d)]\, p(\theta|\boldsymbol{x})\, \mathrm{d}\theta \right\} \\ &= \arg\min_{\theta} \{1 - p(\theta|\boldsymbol{x})\} \\ &= \arg\max_{\theta} \{p(\theta|\boldsymbol{x})\}. \end{aligned}$$

3.3.5 Coût l_p

Les fonctions coût l_p

$$L_p(\theta, d) = |\theta - d|^p,$$

où p est un nombre réel positive, permettent d'effectuer un compromis entre les différentes propriétés des fonctions coût quadratique et absolu, tout en préservant la convexité et la continuité de la dérivée à l'origine avec le choix de p. Pour $p = 1$ on retrouve le coût absolu et pour $p = 2$ le coût quadratique.

3.3.6 Coût Linex

Parmi les fonctions de coût asymétriques on peut citer la fonction

$$C(\theta, d) = b\left[\exp\left[a(\theta - d)\right] - a(\theta - d) - 1\right], \quad a \neq 0, b > 0.$$

Cette fonction, proposé par Varian (1975)[], peut être utilisée lorsqu'on a des raisons pour plus pénaliser les erreurs positives que les erreurs négatives. Plusieurs auteurs ont utilisé cette fonction pour estimation du paramètre de position θ dans un modèle gaussien.

Proposition 6 L'estimateur de Bayes δ de $\theta \in \mathbb{R}$, associé à la loi *a posteriori* $p(\theta|\boldsymbol{x})$ et au coût Linex est

$$\delta(\boldsymbol{x}) = -\frac{1}{a}\log\left(\mathrm{E}\left[\exp\left[-a\theta\right]|\boldsymbol{x}\right]\right).$$

En effet le coût moyen est

$$\bar{C}(\delta) = b\left\{\exp\left[a\delta\right]\mathrm{E}\left[\exp\left[-a\theta\right]|\boldsymbol{x}\right] - a(\delta - \mathrm{E}\left[\theta|\boldsymbol{x}\right] - 1\right\}$$

et l'argument qui minimise cette fonction est

$$\delta = -\frac{1}{a}\log\left(\mathrm{E}\left[\exp\left[-a\theta\right]|\boldsymbol{x}\right]\right),$$

où

$$\mathrm{E}\left[\exp\left[-a\theta\right]|\boldsymbol{x}\right] = \int \exp\left[-a\theta\right] p(\theta|\boldsymbol{x})\,\mathrm{d}\theta.$$

Notons que l'estimateur associé à cette fonction ne dépend pas du coefficient b. Ceci est à comparer au coût quadratique généralisée où l'estimateur associé ne dépend pas des coefficients de pondération. Notons aussi que pour des petites erreurs on a

$$b\left[\exp\left[a\Delta\right] - a(\Delta) - 1\right] \simeq b\Delta^2, \quad \text{si} \quad |a| < 1$$

ce qui signifie que pour $|a| < 1$ et pour des petites erreurs cette fonction coût est approximativement équivalente à une fonction coût quadratique.
Il est évident que dans le cas général l'estimateur bayésien dépend de la fonction coût utilisée. Il existe cependant une famille particulière des lois pour lesquelles les estimateurs bayésiennes sont indépendant de la fonction de coût. Mise en évidence par Rutkin (1978), cette famille est l'ensemble des distributions qui s'écrivent

$$\log p(\theta|\boldsymbol{x}) = t_1(\boldsymbol{x})\exp\left[\alpha\theta\right] + t_2(\boldsymbol{x})\exp\left[-\alpha\theta\right]$$

ce qui peut correspond à

$$\log p(\boldsymbol{x}|\theta) + \log p(\theta) = t_1(\boldsymbol{x})\exp\left[\alpha\theta\right] + t_2(\boldsymbol{x})\exp\left[-\alpha\theta\right] + t_3(\boldsymbol{x})$$

ou encore

$$p(\boldsymbol{x}|\theta) = \frac{B(\boldsymbol{x})}{p(\theta)}\exp\left[t_1(\boldsymbol{x})\exp\left[\alpha\theta\right] + t_2(\boldsymbol{x})\exp\left[-\alpha\theta\right]\right].$$

Notons que ces distributions sont des cas particuliers des lois exponentielles généralisées.
A COMPLETER

3.3.7 Autres fonctions coût

D'autres fonctions coût peuvent être définies et utilisées. Par exemple si on choisit une fonction coût

$$C(\theta, d) = \frac{(\theta - d)^2}{\theta},$$

l'estimateur correspondant est

$$\delta(\boldsymbol{x}) = \frac{1}{\mathrm{E}\left[1/\theta\right]} = \frac{1}{\int \frac{1}{\theta} p(\theta|\boldsymbol{x}) \,\mathrm{d}\theta}.$$

Par contre si on choisit une fonction coût

$$C(\theta, d) = \frac{(\theta - d)^2}{d},$$

l'estimateur correspondant est

$$\delta(\boldsymbol{x}) = \sqrt{\mathrm{E}\left[\theta^2\right]} = \sqrt{\int \theta^2 p(\theta|\boldsymbol{x}) \,\mathrm{d}\theta}.$$

Une autre fonction coût est

$$C(\theta, d) = -\log d(\theta).$$

Le coût moyen est

$$\mathrm{E}\left[C\right] = -\int \log d(\theta)\, p(\theta|\boldsymbol{x}) \,\mathrm{d}\theta$$

qui est minimisé lorsque $d(\theta) = p(\theta|\boldsymbol{x})$. Ainsi l'estimateur correspondant est

$$\delta(\boldsymbol{x}) = p(\theta|\boldsymbol{x}),$$

et coût moyen optimal correspondant est

$$\mathrm{E}\left[C\right]_{\min} = -\int p(\theta|\boldsymbol{x}) \log p(\theta|\boldsymbol{x}) \,\mathrm{d}\theta$$

qui correspond à l'entropie de la loi *a posteriori* $p(\theta|\boldsymbol{x})$.
La différence entre le coût moyen $\mathrm{E}\left[C\right]$ et sa minimum $\mathrm{E}\left[C\right]_{\min}$ est

$$D = -\int p(\theta|\boldsymbol{x}) \log\left(\frac{p(\theta|\boldsymbol{x})}{d(\theta)}\right) \,\mathrm{d}\theta,$$

qui est la distance de Kullback entre $p(\theta|\boldsymbol{x})$ et $d(\theta)$.

3.3.8 Cas d'un vecteur de paramètres

L'extension au cas d'un vecteur des paramètres $\boldsymbol{\theta}$ se fait on définissant une fonction coûts $C(\boldsymbol{\theta}, \boldsymbol{d})$, en calculant le coût moyen

$$\bar{C}(\boldsymbol{d}) = \mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{x}}\left\{C(\boldsymbol{\theta}, \boldsymbol{d})\right\} = \int C(\boldsymbol{\theta}, \boldsymbol{d}) p(\boldsymbol{\theta}|\boldsymbol{x}) \,\mathrm{d}\boldsymbol{\theta},$$

et en définissant l'estimateur bayésien par

$$\boldsymbol{\delta}(\boldsymbol{x}) = \arg\min_{\boldsymbol{d}} \left\{\bar{C}(\boldsymbol{d})\right\}.$$

Les différentes fonctions coûts usuelles sont

- coût quadratique

$$C(\boldsymbol{\theta}, \boldsymbol{d}) = [\boldsymbol{\theta} - \boldsymbol{d}]^t \boldsymbol{Q} [\boldsymbol{\theta} - \boldsymbol{d}] = \sum_i \sum_j q_{i,j} (\theta_i - d_i)(\theta_j - d_j).$$

  L'estimateur correspondant est

$$\boldsymbol{\delta}(\boldsymbol{x}) = \mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{x}} \{\boldsymbol{\theta}\} = \int \boldsymbol{\theta} p(\boldsymbol{\theta}|\boldsymbol{x}) \, \mathrm{d}\boldsymbol{\theta}.$$

  Cet estimateur est appelé *moyenne a posteriori* (MP).

- coût Dirac

$$C(\boldsymbol{\theta}, \boldsymbol{d}) = 1 - u(\boldsymbol{\theta} - \boldsymbol{d}),$$

  où $u(\boldsymbol{\theta} - \boldsymbol{d})$ est la distribution de Dirac. L'estimateur correspondant est

$$\boldsymbol{\delta}(\boldsymbol{x}) = \arg\max_{\boldsymbol{\theta}} \{p(\boldsymbol{\theta}|\boldsymbol{x})\}$$

  Cet estimateur est appelé *maximum a posteriori* (MAP).

- coût absolue

$$C(\boldsymbol{\theta}, \boldsymbol{d}) = \sum_i a_i |\theta_i - d_i|$$

  L'estimateur correspondant est

$$\delta_i(\boldsymbol{x}) = \text{Méd} \left[p(\theta_i|\boldsymbol{x}) \right],$$

  où $p(\theta_i|\boldsymbol{x})$ est la loi marginale.

- coût Dirac marginal

$$C(\boldsymbol{\theta}, \boldsymbol{d}) = \sum_i u(\theta_i - d_i)$$

$$\delta_i(\boldsymbol{x}) = \arg\max_{\theta_i} \{p(\theta_i|\boldsymbol{x})\}$$

  Cet estimateur est appelé *maximum a posteriori marginale* (MPM).

3.3.9 Lien entre les estimateurs de Bayes et caractéristiques de la loi *a posteriori*

Comme nous avons vue, il y a un lien étroit entre les différentes caractéristiques d'une loi et les estimateurs bayésiens avec les fonctions coût usuelles. La proposition suivante résume un certain nombre de ces liens.
Étant donnée une fonction densité de probabilité $p(x)$, considérons l'expression suivante :

$$J_p(m) = \left| \int_{-\infty}^{+\infty} |x - m|^p \, p(x) \, \mathrm{d}x \right|^{\frac{1}{p}}.$$

Pour p donnée, la valeure m_p qui minimise $J_p(m)$ est appellée le *centre* de $p(x)$ au sense de la norme l_p, et la valeure minimum σ_p de $J_p(m)$ est appellée la *dispersion* de $p(x)$ au sense de la norme l_p :

$$m_p = \arg\min \{J_p(m)\}$$

$$\sigma_p = J_p(m_p)$$

Pour p=1, la valeure m_1 est appellée la *médiane*, m_2 la *moyenne* et m_∞ le *mi–support* de $p(x)$.
Pour p=2, la valeure σ_1 est appellée l'*écart-moyen*, σ_2 la *variance* et σ_∞ le *demi–support* de $p(x)$.

– médiane (argmin de la norme l_1):

$$m_1 = \arg\min \left\{ \int_{-\infty}^{+\infty} |x - m_1| \, p(x) \, \mathrm{d}x \right\} \Longrightarrow \int_{-\infty}^{m_1} p(x) \, \mathrm{d}x = \int_{m_1}^{+\infty} p(x) \, \mathrm{d}x = \frac{1}{2}$$

– moyenne (argmin de la norme l_2):

$$m_2 = \arg\min \left\{ \int_{-\infty}^{+\infty} |x - m_2|^2 \, p(x) \, \mathrm{d}x \right\} \Longrightarrow m_2 = \int_{-\infty}^{+\infty} x \, p(x) \, \mathrm{d}x$$

– mi–support (argmin de la norme l_∞):

$$m_\infty = \arg\min \left\{ \lim_{p \mapsto \infty} \int_{-\infty}^{+\infty} |x - m_\infty|^p \, p(x) \, \mathrm{d}x \right\} \Longrightarrow m_\infty = \frac{x_{\mathrm{sup}} + x_{\mathrm{inf}}}{2}$$

– écart-moyen (minimum de la norme l_1):

$$\sigma_1 = \int_{-\infty}^{+\infty} |x - m_1| \, p(x) \, \mathrm{d}x \Longrightarrow \sigma_1 = \int_{m_1}^{\infty} x \, p(x) \, \mathrm{d}x - \int_{-\infty}^{m_1} x \, p(x) \, \mathrm{d}x$$

– variance (minimum de la norme l_2):

$$m_2 = \int_{-\infty}^{+\infty} |x - m_2|^2 \, p(x) \, \mathrm{d}x \Longrightarrow \sigma_2 = \int_{-\infty}^{+\infty} (x - m_2)^2 \, p(x) \, \mathrm{d}x = \int_{-\infty}^{+\infty} x^2 \, p(x) \, \mathrm{d}x - m_2^2$$

– demi–support (minimum de la norme l_∞):

$$\sigma_\infty = \lim_{p \mapsto \infty} \left| \int_{-\infty}^{+\infty} |x - m_\infty|^p \, p(x) \, \mathrm{d}x \right|^{\frac{1}{p}} \Longrightarrow \sigma_\infty = \frac{x_{\mathrm{sup}} - x_{\mathrm{inf}}}{2}$$