Remarque 4 (Identifiabilité (e.g. voir [Mon97, Chap 7]))

Un problème classique en statistique est liée à l'hypothèse d'injectivité de l'application $\theta \mapsto \mathbb{P}_\theta$, autrement dit la condition qui garantit que deux paramètres différents correspondent à deux lois distinctes. Il s'agit en général d'une condition difficile à vérifier.

Exemple. un n-échantillon $X_1, \ldots, X_n$ d'une loi $\mathcal{N}(\theta, \sigma^2)$ avec comme seule observations $Y_i = |X_i|, i = 1, \ldots, n$: alors θ n'est pas identifiable.

- On peut définir une notion d'identifiabilité locale, à savoir la valeur θ_0 est localement identifiable si dans un voisinage de θ_0, $\mathbb{P}_\theta \neq \mathbb{P}_{\theta_0}$. Et le modèle est dit localement identifiable si la propriété précédente est ok pour toute valeur de θ_0.
- Alors on peut montrer que si la matrice d'information de Fisher $I(\theta)$ au voisinage de θ_0 est définie positive alors θ_0 est localement identifiable. Donc si $I(\theta)$ est définie positive pour tout θ le modèle est localement identifiable. Mais ceci n'entraîne pas l'identifiabilité en général **sauf par exemple pour les modèles exponentiels**.

Statistique associée à un échantillon

Définition 10

*On appelle **statistique** associée à un n-échantillon du modèle $(E_X, \mathscr{E}, (\mathbb{P}_\theta)_{\theta\in\Theta})$, une fonction mesurable t_n sur ${E_X}^n$ à valeurs dans $t_n(E_X) \subset \mathbb{R}^p$ et indépendante de θ. Notons que $t_n(x_1, \ldots, x_n)$ est une réalisation de la v.a. $T_n := t_n(X_1, \ldots, X_n)$. Par abus, le terme statistique sera donné aux deux quantités t_n et T_n.*

Statistiques empiriques

Observation unidimensionnelle ($E_X \subset \mathbb{R}$)

1. Moyenne empirique $t_n(x_1, \ldots, x_n) = \overline{x}_n := (1/n)\sum_{i=1}^n x_i$: $\overline{X}_n := (1/n)\sum_{i=1}^n X_i = t_n(X_1, \ldots, X_n)$
2. Variance empirique : $t_n(x_1, \ldots, x_n) = s_n^2 := (1/n)\sum_{i=1}^n (x_i - \overline{x}_n)^2 = \overline{x^2}_n - (\overline{x}_n)^2$ $t_n(X_1, \ldots, X_n) = S_n^2 := (1/n)\sum_{i=1}^n (X_i - \overline{X}_n)^2 = \overline{X^2}_n - (\overline{X}_n)^2$

Observation multidimensionnelle ($E_X \subset \mathbb{R}^p$)

1. Moyenne empirique : $\overline{X}_n = \left(\overline{X(k)}_n\right)_{k=1}^p := \left(\frac{1}{n}\sum_{i=1}^n X_i(k)\right)_{k=1}^p$
2. Covariance empirique entre $X(k)$ et $X(l)$: $S(k,l)_n := \frac{1}{n}\sum_{i=1}^n (X_i(k) - \overline{X(k)}_n)(X_i(l) - \overline{X(l)}_n)$

Application du Théorème de Cochran :

Proposition 9 (Moments empiriques)

Soit $(X_1, \ldots, X_n)$ un n-échantillon d'une loi gaussienne (m, σ^2) alors

1. $\overline{X}_n$ *et* S_n^2 *sont indépendants*
2. $\sqrt{n}\left[\frac{\overline{X}_n - m}{\sigma}\right] \sim \mathcal{N}(0,1)$, $\frac{1}{\sigma^2}\sum_{i=1}^n (X_i - m)^2 \sim \chi_n^2$
3. $\frac{nS_n^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \overline{X}_n)^2 \sim \chi_{n-1}^2$, $\sqrt{n-1}\left[\frac{\overline{X}_n - m}{S_n}\right] \sim st_{n-1}$

Remarque 5

- Il existe une réciproque à la propriété 1. À savoir, si les deux statistiques pour un n-échantillon sont indépendantes alors la loi commune est nécessairement gaussienne.
- Notons que seules les v.a. du 1. sont des statistiques.
- La v.a. $\widetilde{S}_n := \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$ sera également largement utilisée par la suite. Notons qu'alors

$$\frac{(n-1)\widetilde{S}_n^2}{\sigma^2} \sim \chi_{n-1}^2, \quad \sqrt{n}\left[\frac{\overline{X}_n - m}{\widetilde{S}_n}\right] \sim st_{n-1}.$$

Statistique libre

Définition 11 (Statistique libre)

Pour un modèle $(E_X, \mathscr{E}, (\mathbb{P}_\theta)_{\theta\in\Theta})$, une statistique sur ce modèle est dite libre si sa loi ne dépend pas du paramètre θ.

Statistiques empiriques pour le modèle gaussien $(\mathcal{N}(\theta,1))_{\theta\in\mathbb{R}}$

La statistique nS_n^2 ou $(n-1)\widetilde{S}_n^2$

Proposition 10

Pour une statistique libre, on a $I^t(\theta) = 0$ pour tout $\theta \in \Theta$.
Si le modèle image par la statistique est homogène, vérifie les conditions de régularité (CD1) et si $\nabla L^t(\theta, \cdot)$ est continue en θ alors la réciproque est également vraie.

Statistique exhaustive

À l'opposé d'une statistique libre qui ne contient aucune information sur le paramètre inconnu, nous allons introduire la notion de statistique exhaustive qui « résume » l'information sur le paramètre contenue dans un échantillon

Définition 12 (Statistique exhaustive)

Pour un modèle statistique $(E_X, \mathscr{E}, (\mathbb{P}_\theta)_{\theta\in\Theta})$, la statistique t est dite exhaustive si la loi conditionnelle de X sachant $t(X)$ est indépendante de θ.

Modèle d'un n-échantillon d'une loi de Bernoulli

On considère le modèle statistique $\left(\{0,1\}^n, (\mathrm{Ber}(\theta))_{\theta\in]0,1[}\right)$. Les observations x sont des séquences binaires $(x_1,\ldots,x_n)$ de longueur n. Intuitivement, l'information sur le paramètre θ (vue comme un proportion de succès dans la modélisation) contenue dans une telle séquence n'est liée qu'au nombre total de 1 de la séquence $t(x)=\sum_i x_i$. La loi de $X=(X_1,\ldots,X_n)$ conditionnelle à la donnée de la valeur de $t(X)=\sum_i X_i = y$

$$\mathbb{P}_\theta\{X=x \mid t(X)=y\} = \begin{cases} 0 & \text{si } t(x)\neq y \\ \frac{1}{C_n^y} & \text{si } t(x)=y \end{cases}$$

est indépendante du paramètre θ.

Théorème 1 (Critère de factorisation)

Pour un modèle statistique $(E_X, \mathscr{E}, (\mathbb{P}_\theta)_{\theta\in\Theta})$, la statistique t à valeurs dans un espace $\mathcal{Y}$ est exhaustive ssi il existe une application g de $\Theta\times\mathcal{Y}$ dans $\mathbb{R}_+$ et une application h de E_X dans $\mathbb{R}_+$ telles que la vraisemblance L s'écrive

$$\forall x\in E_X, \forall\theta\in\Theta, \quad L(\theta,x) = g(\theta,t(x))\,h(x) \tag{6}$$

Remarque 6

Un choix approprié de fonction $g(\theta,\cdot)$ permet d'interpréter ce terme comme la vraisemblance du modèle image et $h(\cdot)$ comme un rapport de vraisemblance.

n-échantillon Bernoulli

$t(x)=\sum_i x_i$ est une statistique exhaustive pour la moyenne inconnue θ.

n-échantillon gaussien de variance connue

$\mathbb{P}_\theta = \mathcal{N}(\theta,\sigma^2)$ avec $\Theta=\mathbb{R}$ et σ^2 connue :

$$\begin{aligned} L(\theta,x) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^n (x_i-\overline{x})^2 + n(\overline{x}_n-\theta)^2\right)\right) \\ &= \exp\left(-\frac{n(\overline{x}_n-\theta)^2}{2\sigma^2}\right) \times \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\overline{x}_n)^2\right) \end{aligned}$$

$\overline{x}_n$ (ou $\overline{X}_n$) est une statistique exhaustive pour la moyenne inconnue θ.

n-échantillon gaussien

La statistique $t(x)=(\overline{x}_n, s_n^2)$ est exhaustive pour $\theta=(m,\sigma^2)$.

Famille exponentielle

Pour une famille exponentielle (voir (1))

$$L(\theta,x) = \beta(\theta)\xi(x)\exp\left(\langle t(x),\alpha(\theta)\rangle\right)$$

la statistique naturelle $t(\cdot)$ est exhaustive.

Proposition 11 (Information de Fisher et exhaustivité)

Dans un modèle homogène, si t est exhaustive alors $I^t(\theta)=I(\theta)$ pour tout $\theta\in\Theta$.

Il est possible de prouver une réciproque de ce dernier résultat sous des hypothèses de continuité des gradients des vraisemblances associés aux modèles initial et image (voir e.g. [Mon97, p 70]).

Théorème 2

Pour deux statistiques s et t sur un modèle $(E_X, \mathscr{E}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ homogène. Si s et t sont indépendantes (c'est à dire indépendantes quelque soit θ) et s exhaustive alors $I^t(\theta)=0$, pour tout θ, ce qui implique que t est libre sous les conditions de validité de la réciproque de la Prop. 10.

La définition suivante met en ouvre l'idée d'une statistique exhaustive réduisant au maximum l'ensemble des données pertinentes en termes d'information sur θ.

Statistique exhaustive minimale

Pour un modèle statistique $(E_X, \mathscr{E}, (\mathbb{P}_\theta)_{\theta\in\Theta})$, une statistique exhaustive est dite minimale si pour toute statistique exhaustive t', la statistique t est une fonction de t' ($t = \varphi \circ t'$ pour une certaine application φ)

Théorème 3 (Lehmann et Scheffé)

Soit t une statistique sur un modèle statistique $(E_X, \mathscr{E}, (\mathbb{P}_\theta)_{\theta\in\Theta})$, alors si on a l'équivalence

$$t(x) = t(y) \iff \frac{L(\theta,x)}{L(\theta,y)} \text{est une fonction indépendante de } \theta$$

alors t est une statistique exhaustive minimale.

Statistique complète

Définition 13 (Modèle et statistique complets)

Un modèle $(E_X, \mathscr{E}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ est dit complet si pour toute fonction numérique g telle que g est $(\mathbb{P}_\theta)_{\theta\in\Theta}$ intégrable, alors

$$(\forall\theta \in \Theta, \quad \mathbb{E}_\theta[g(X)] = 0) \Longrightarrow \forall\theta, \quad g = 0 \; \mathbb{P}_\theta \; ps$$

Une statistique t sur $(E_X, \mathscr{E}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ est dite complète si le modèle image par t est complet, i.e. pour toute fonction numérique g telles que g est $(\mathbb{P}_\theta^t)_{\theta\in\Theta}$ intégrable on a

$$\left(\forall\theta \in \Theta, \quad \int g(y) d\mathbb{P}_\theta^t(y) = 0\right) \Longrightarrow \forall\theta, \quad g = 0 \; \mathbb{P}_\theta^t \; ps$$

ou encore par le théorème de transfert

$$(\forall\theta \in \Theta, \quad \mathbb{E}_\theta[g(t(X))] = 0) \Longrightarrow \forall\theta, \quad g \circ t = 0 \; \mathbb{P}_\theta \; ps$$

Notons que toute statistique définie sur un modèle complet est complète.

Remarque 7 (Mesure de Lebesgue)

Si g est mesurable sur $\mathbb{R}$ d'intégrale nulle sur tout intervalle compact alors $g = 0$ presque partout.

Modèle binomial

$t(x) = x$ est une statistique complète sur $(\{0,\ldots,n\}, (\mathrm{Bin}(n,\theta))_{\theta\in]0,1[})$ (i.e. le modèle est complet)

Famille exponentielle

On peut montrer (voir e.g. [Mon97, Th 4, p 61]) que pour une famille exponentielle de vraisemblance (1), si $\alpha(\Theta) \subset \mathbb{R}^p$ est d'intérieur non vide alors la statistique naturelle t est complète.

Théorème 4

Pour deux statistiques s et t sur $(E_X, \mathscr{E}, (\mathbb{P}_\theta)_{\theta\in\Theta})$. Si s est exhaustive et complète et si t est libre alors s et t sont indépendantes (c'est à dire indépendantes quelque soit θ).

Estimation d'un paramètre

Ampoule

On a mesuré la durée de vie en h de 10 ampoules identiques :

$$x_1 = 91.6, \; x_2 = 35.7, \; x_3 = 251.3, \; x_4 = 24.3, \; x_5 = 5.4,$$
$$x_6 = 67.3, \; x_7 = 170.9, \; x_8 = 9.5, \; x_9 = 118.4, \; x_{10} = 57.1$$

Modèle statistique : la durée de vie d'une ampoule est supposée être la réalisation d'une variable aléatoire de loi exponentielle de paramètre $\theta \in]0, +\infty[$ et les 10 données sont les valeurs observées d'un 10-échantillon de la loi $\mathrm{Exp}(\theta)$.

On suppose que la famille des lois exponentielles est adaptée mais la valeur de θ reste à déterminer

L'estimation de θ

Objectif : affecter une valeur $\widehat{\theta}$ au paramètre θ, i.e. choisir une loi $\mathbb{P}_{\widehat{\theta}}$ dans $\mathcal{P}$

1. **Paradigme fréquentiste :** affecter une valeur uniquement sur la base d'information « objective ». L'information objective est portée par les seules observations ou données
2. **Paradigme bayesien :** autoriser l'introduction d'information dite *a priori* sur le paramètre qui se traduit généralement par la spécification par l'analyste d'une loi dite *a priori* de θ. Puis on met à jour l'information sur la loi de θ au vu des observations : calcul d'une loi dite *a posteriori* de θ à l'aide de la formule de Bayes. [Rob06]
   Cette approche est devenu centrale dans certains domaines comme la classification [Bis06, HTF01].

On se concentre dans ce cours sur la première approche, dite aussi « approche classique » : **dans toute la partie estimation, le modèle statistique sera celui d'un n-échantillon**

C. Bishop.
Pattern Recognition and Machine Learning.
Springer, 2006.

T. Hastie, R. Tibshirani et J. H. Friedman.
The Elements of Statistical Learning.
Springer, 2001.

C. Robert.
Le Choix Bayésien : Principes et Pratique
Springer, 2006

Ampoule (suite)

Les données :

$$x_1 = 91.6,\ x_2 = 35.7,\ x_3 = 251.3,\ x_4 = 24.3,\ x_5 = 5.4,$$
$$x_6 = 67.3,\ x_7 = 170.9,\ x_8 = 9.5,\ x_9 = 118.4,\ x_{10} = 57.1$$

Le paramètre θ est égal à l'inverse de la valeur espérée (ou moyenne) $\mathbb{E}_\theta[X]$ de la durée de vie d'où une tentation naturelle

$$\widehat{\theta}(x_1, \ldots, x_{10}) = \frac{1}{\overline{x}_{10}} = \frac{10}{\sum_{i=1}^{10} x_i} = \frac{1}{83.15} \approx 0,012$$

et la loi exponentielle choisie est donc $\mathrm{Exp}(0,012)$

↪ le modèle est complètement spécifié et toutes les prévisions sont accessibles

- Mais la variabilité sur les valeurs des données suggère d'étudier les fluctuations de la v.a. $1/\overline{X}_{10} = 10/\sum_{i=1}^{10} X_i$. Leur contrôle permettrait de quantifier la qualité de l'estimation obtenue sur un jeu de données spécifique pour la population de toutes les ampoules de même type.

1. Méthodes pour estimer un paramètre
2. Définir des mesures de la qualité de l'estimation obtenue.

Définition 14 (Estimation / Estimateur)

On considère un modèle statistique de n-échantillon.

- *On appelle* ***estimation*** *$\widehat{\theta}$ de θ toute statistique sur le modèle à valeurs au moins dans Θ. Autrement dit, il s'agit d'une fonction mesurable $\widehat{\theta}(x_1, \ldots, x_n)$ des n observations et $\widehat{\theta}(x_1, \ldots, x_n)$ représente la valeur attribuée au paramètre θ.*
- *La v.a. $\widehat{\Theta} = \widehat{\theta}(X_1, \ldots, X_n)$ est appelée un* ***estimateur*** *de θ. La valeur observée de cette v.a. fournit une estimation de θ.*

Statistiques empiriques

Les statistiques empiriques $\overline{X}_n$ et S_n^2 introduites dans l'Exemple 1 sont des estimateurs naturels de l'espérance et la variance d'un loi donnée (cf Stat. Descriptive).

La méthode des moments I

- Il s'agit de la méthode la plus naturelle lorsque le paramètre à estimer peut s'interpréter comme la valeur d'une certaine espérance sous $\mathbb{P}_\theta$.

  Par exemple, on estime l'espérance de la loi par la moyenne empirique, la variance par la variance empirique, la covariance par la covariance empirique, ...
- **Principe général :** définir l'estimation de θ comme la valeur du paramètre pour laquelle un (ou plusieurs) moment théorique est le plus proche possible du (ou des) moment empirique correspondant.

La méthode des moments II

1. Soit $\mathbb{E}_\theta[\varphi(X)]$ (par exemple $\varphi(x) = x, \varphi(x) = x^2, \varphi(x) = \exp(x), \ldots$)
2. Soit $m(\theta)$ son expression en fonction du paramètre θ
3. Soit $\overline{m}_n$ le moment empirique correspondant
$$\overline{m}_n = \frac{1}{n}\sum_{i=1}^{n}\varphi(x_i)$$
4. L'estimation des moments de θ est la solution de l'équation
$$m(\theta) = \overline{m}_n. \tag{7}$$
   - Si $m(\theta)$ est inversible alors l'estimation des moments est
   $$\widehat{\theta} = m^{-1}(\overline{m}_n)$$
   - Si l'équation (7) n'admet aucune solution alors on peut par exemple définir l'estimation des moments comme la valeur de θ qui minimise $|m(\theta) - \overline{m}_n|$.

La méthode des moments III

1. loi de Bernoulli $\mathrm{Ber}(\theta)$: $\mathbb{E}_\theta[X] = \theta$ donc l'estimation des moments $\widehat{\theta}$ de θ est $\overline{x}_n$.
2. loi exponentielle $\mathrm{Exp}(\theta)$: $\mathbb{E}_\theta[X] = 1/\theta$ donc l'estimation des moments $\widehat{\theta}$ de θ est $1/\overline{x}_n$
3. loi normale $\mathcal{N}(\theta)$ avec $\theta = (m, \sigma^2)$: $\mathbb{E}_\theta[X] = m$ et $\mathbb{E}_\theta[X^2] = \sigma^2 + m^2$ donc on résout $(\mathbb{E}_\theta[X], \mathbb{E}_\theta[X^2]) = (m, \sigma^2 + m^2)$ et on obtient
$$\widehat{m} = \overline{x}_n \qquad \widehat{\sigma^2} = \overline{x^2}_n - \overline{x}_n{}^2 = \frac{1}{n}\sum_i (x_i - \overline{x}_n)^2 = s_n^2$$

Cela donne une estimation des moments pour la variance d'une loi quelconque (avec une moyenne inconnue)

La méthode des moments IV

- Des avantages :
  - Le principe de la méthode ne s'appuie pas sur la propriété d'indépendance de l'échantillon. Elle peut donc être employée même si l'indépendance des observations n'est pas réalisée.
  - Simplicité et support intuitif du calcul
- Des désavantages :
  - Le choix du (des) moment(s) est pas toujours simple
  - En général, la méthode des moments ne permet d'exploiter pleinement les caractéristiques du modèle. Par conséquent, l'estimation des moments sera en général moins précise que d'autres méthodes comme celle du maximum de vraisemblance.

La méthode du maximum de vraisemblance I

- **Principe général :** l'estimation du maximum de vraisemblance (EMV) de θ est la valeur du paramètre (i.e. la loi $\mathbb{P}_\theta$ de $\mathcal{P}$) pour laquelle la vraisemblance d'observation de $(x_1, \ldots, x_n)$ est maximale
  Rmq : cette méthode ne s'applique qu'aux modèles dominés.
- Par exemple dans le cas de lois discrètes, la vraisemblance de l'échantillon est donnée par

$$L_n(\theta, x) = \prod_{i=1}^{n} \mathbb{P}_\theta\{x_i\} = \mathbb{P}_\theta\{(x_1, \ldots, x_n)\}$$

Autrement dit, la vraisemblance est exactement la probabilité d'observer $(x_1, \ldots, x_n)$ sous $\mathbb{P}_\theta$. Dans ce cas, l'EMV correspond à la valeur de θ (i.e. la loi $\mathbb{P}_\theta$) pour laquelle cette probabilité d'observer « les données » sera maximale.

La méthode du maximum de vraisemblance II

Définition 15 (EMV)

Soit $(E_X, \mathscr{E}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ un modèle statistique (dominé). Une estimation de maximum de vraisemblance (EMV) est une statistique $\widehat{\theta}(\cdot)$ sur $(E_X, \mathscr{E}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ vérifiant

$$L(\widehat{\theta}, x) = \sup_{\theta\in\Theta} L(\theta, x) \qquad \text{pour } \mu\text{-tout } x \in E_X \tag{8}$$

ou ce qui s'écrit parfois sous la forme

$$\widehat{\theta}(x) = \arg\max_{\theta\in\Theta} L(\theta, x) \qquad \text{pour } \mu\text{-tout } x \in E_X$$

Remarque 8 (Maximisation de la ln-vraisemblance / Équation de vraisemblance)

1. La vraisemblance dans le cas des n-échantillons est un produit de fonctions. Le passage au $\ln$ permet de linéariser de linéariser la fonction objectif. Comme la fonction $\ln$ est strictement croissante, la maximisation de la vraisemblance est équivalente à celle de $\ln$-vraisemblance
2. Si $\theta \mapsto L(\theta, x)$ est suffisamment régulière alors l'EMV satisfait nécessairement l'équation dite de vraisemblance :

$$\nabla L(\theta, x) = 0 \quad \text{ou} \quad \nabla \ln L(\theta, x) = 0 \tag{9}$$

La méthode du maximum de vraisemblance III

Des exemples

1. loi de Bernoulli $\mathrm{Ber}(\theta)$: l'estimation EMV de θ est $\overline{x}_n$.
2. loi exponentielle $\mathrm{Exp}(\theta)$: l'estimation EMV de θ est $1/\overline{x}_n$
3. loi uniforme $\mathrm{Unif}[0, \theta]$: l'estimation EMV de θ est $x_{(n)} = \max(x_1, \ldots, x_n)$
4. modèle de régression simple : $(Y_1, \ldots, Y_n)$ est un vecteur gaussien de moyenne $(ax_i + b)_{i=1}^n$ et de matrice de covariance $\sigma^2 I_n$ et Exo 1, feuille 7 TD de Stat. Desc. que les estimations EMV des coefficients de la régression sont identiques à celles des MC :

$$\widehat{a}(y_1, \ldots, y_n) = \frac{\sum_i (x_i - \overline{x}_n)(y_i - \overline{y}_n)}{\sum_i (x_i - \overline{x})^2}$$
$$\widehat{b}(y_1, \ldots, y_n) = \overline{y}_n - \widehat{a}(y_1, \ldots, y_n)\,\overline{x}_n$$

mais pour l'estimateur de la variance

$$\widehat{\sigma^2}_{\mathrm{EMV}} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - (\widehat{a}x_i + \widehat{b})\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\widehat{e}_i^2 \qquad \widehat{\sigma^2} = \frac{1}{n-2}\sum_{i=1}^{n}\widehat{e}_i^2$$

La méthode du maximum de vraisemblance IV

- Des avantages :
  - Le principe de la méthode est attractif
  - **Propriété d'invariance** : il arrive régulièrement que l'on veuille estimer non pas θ mais une fonction $g(\theta)$. Alors pour des transformations bijectives (et dans la cas contraire, on prend la relation comme définition)

$$\widehat{g(\theta)} = g(\widehat{\theta}). \tag{10}$$

  - S'il existe une statistique exhaustive t alors l'EMV est une fonction de cette statistique.
- Des désavantages :
  - Notons qu'il s'agit d'un problème d'optimisation global de la fonction $\theta \mapsto L(\theta, x)$ avec tous les problèmes théoriques et pratiques associés :
    - il n'y a aucune garantie que (8) ou (9) admette une solution ou qu'il existe une unique solution.
    - la résolution explicite est souvent difficile voire impossible : recours à des méthodes numériques d'optimisation ou de résolution d'équations
      Pbs d'initialisation, de convergence et de vitesse de convergence des algorithmes.