Cours 5 : ESTIMATION PONCTUELLE

A- Généralités
B- Précision d'un estimateur
C- Exhaustivité
D- information
E-estimateur sans biais de variance minimale, estimateur efficace
F- Quelques méthode s d'estimation

A- ESTIMATION PONCTUELLE: GÉNÉRALITÉS

A- 1 Définition

- On s'intéresse à la caractéristique X d'une population (éventuellement à un vecteur de caractéristiques), dont la loi dépend d'un paramètre inconnu

$$\theta \in \vartheta \subset \mathbb{R}^p, p \geq 1.$$

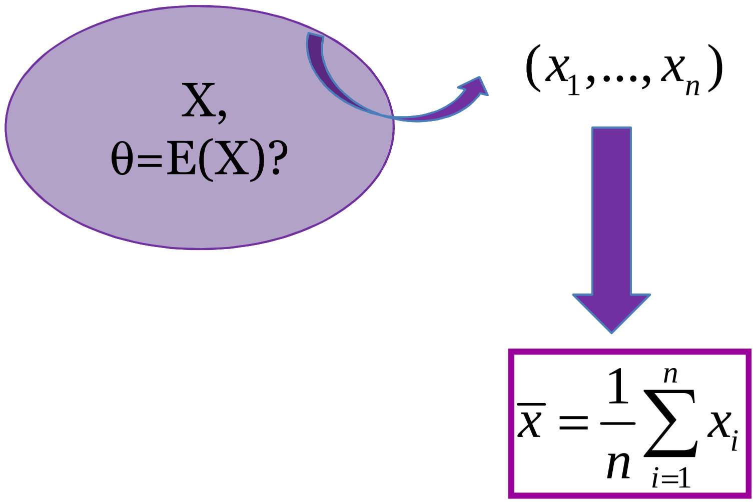
- On note $f_\theta(x)$ la densité de la loi de X au point x (resp. la loi $P_\theta(X=x)$ de X au point x) si X est continue (resp. si X est discrète).

- On dispose d'un sondage de taille n de la population (l'observation de X sur n individus) , noté $(x_1,\ldots,x_n)$. On note $(X_1,\ldots,X_n)$ l'échantillon aléatoire associé à ce sondage (il s'agit d'un vecteur aléatoire dont une réalisation particulière est $(x_1,\ldots,x_n)$.

A-1 Définition

- ✓ Estimer le paramètre θ consiste à donner une valeur approchée à ce paramètre à partir d'un sondage de la population.

Ex : P



$\overline{x}$ est une approximation ou **estimation** de θ.

A-1 Définition

✓ **Estimateur et estimation :**

P

X,

θ=E(X)?

$$\boxed{\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i} \Longleftarrow (x_1,\ldots,x_n) \qquad (x'_1,\ldots,x'_n) \Longrightarrow \boxed{\overline{x'} = \frac{1}{n}\sum_{i=1}^{n} x'_i}$$

•

$(x_1,\ldots,x_n)$ $(x'_1,\ldots,x'_n)$ sont deux réalisations de l'échantillon aléatoire $(X_1,\ldots,X_n)$

- Les deux **estimations** $\overline{x}$ et $\overline{x'}$ de θ sont deux réalisations de la statistique $\boxed{\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i}$ appelée **estimateur** de θ.

A-1 Définition

✓ **Formalisation** :

Soit $\theta \in \vartheta \subset \mathbb{R}^p, p \geq 1$ le paramètre inconnu dont dépend la loi de X

- Un **estimateur** Θ_n de θ est une statistique de l'échantillon aléatoire :

$$\boxed{\Theta_n = h(X_1,\ldots,X_n), \quad h : \mathbb{R}^n \to \mathbb{R}^p,\ p \geq 1}$$

telle que pour chaque réalisation $(x_1,\ldots,x_n)$ de l'échantillon aléatoire, la valeur $\hat{\theta}_n = h(x_1,\ldots x_n)$ prise par Θ_n approche θ.

- $\hat{\theta}_n$ s'appelle une **estimation** de θ . C'est une réalisation particulière de l'estimateur Θ_n .

A-2 Exemple d'estimateurs

- ✓ **Estimateur de l'espérance E(X) de X : La moyenne empirique**

$$\boxed{\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i}$$

Pour une réalisation donnée $(x_1, \ldots, x_n)$ de l'échantillon aléatoire,
$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$ est l'estimation de E(X) associée à ce jeu de données.

Propriét és :

$$E(\bar{X}_n) = E(X)$$

$$V(\bar{X}_n) = \frac{V(X)}{n}$$

A-2 Exemple d'estimateurs

- ✓ **Estimateurs de la variance σ² et de l'écart-type σ de X**

- ✓ lorsque E(X)=m est connue :

$$\boxed{T_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2}$$ est un estimateur de σ²

$$\boxed{T_n = \sqrt{T_n^2}}$$ est un estimateur de σ

Pour une réalisation donnée de l'échantillon aléatoire, $t_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2$ et t_n

Sont les estimations associées .

Propriétés (cf TD) :

$$E(T_n^2) = \sigma^2$$

$$V(T_n^2) = \frac{\mu^4 - \sigma^4}{n}$$

A-2 Exemple d'estimateurs

✓ Cas général

➢ La variance et l'écart-type empirique s

$$\boxed{S_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2}$$ est un estimateur de σ²

$$\boxed{S_n = \sqrt{S_n^2}}$$ est un estimateur de σ

Pour une réalisation donnée de l'échantillon aléatoire, $s_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$ et

$s_n = \sqrt{s_n^2}$ sont les estimations associées (on les note encore $\widehat{\sigma^2}$ et $\hat{\sigma}$

Propriétés:

$$E(S_n^2) = \frac{n-1}{n}\sigma^2$$

$$V(S_n^2) = \frac{n-1}{n^3}\left((n-1)\mu^4 - (n-3)\sigma^4\right)$$

A-2 Exemple d'estimateurs

- La variance et l'écart-type empiriques corrigés:

$$\boxed{S_n^{*2} = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2}$$ est un estimateur de σ²

$$\boxed{S_n^{*} = \sqrt{S_n^{*2}}}$$ est un estimateur de de σ

Propriétés :

$$E(S_n^{2*}) = \sigma^2$$

$$V(S_n^{2*}) = \frac{1}{n}\left(\mu^4 - \frac{(n-3)}{(n-1)}\sigma^4\right)$$

A-2 Exemple d'estimateurs

- ✓ **Estimateur de la fonction de répartition F(x) : La fonction de répartition empirique:**

$$\boxed{F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{X_i < x}}$$

est un estimateur de F(x) en tout point x.

Pour une réalisation donnée $(x_1, \ldots, x_n)$ de l'échantillon aléatoire,

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{x_i < x}$$

est l'estimation de F(x) associée à ce jeu de données.

Propri étés :

$$E(F_n(x)) = F(x)$$

$$V(F_n(x)) = \frac{F(x)(1 - F(x))}{n}$$

A-2 Exemple d'estimateurs

Soient $(X_1,\dots,X_n)$ et $(Y_1,\dots,Y_n)$ les échantillons aléatoires associés aux variables aléatoires X et Y de P.

- ✓ **Estimateurs de la covariance cov(X,Y) entre deux v.a. X et Y**

- ✓ Lorsque m1=E(X) et m2=E(Y) sont connues

$$\boxed{T_n(X,Y)=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-m_1\right)\left(Y_i-m_2\right)}$$

est un estimateur de cov(X,Y).

Pour une réalisation particulière des échantillons aléatoires,

$$t_n(x,y)=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-m_1\right)\left(y_i-m_2\right)$$

est l'estimation associée

A-2 Exemple d'estimateurs

✓ Cas général

➢ La covariance empirique :

$$\boxed{S_n(X,Y)=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)\left(Y_i-\overline{Y}\right)}$$

est un estimateur de cov(X,Y).

Pour une réalisation particulière des échantillons aléatoires,

$$s_n(X,Y)=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right)$$

(encore notée $\widehat{cov}(X,Y)$)

est l'estimation associée

A-2 Exemple d'estimateurs

- La covariance empirique corrigée :

$$\boxed{S_n^*(X,Y)=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)}$$ est un estimateur de cov(X,Y).

Propriétés :

$$E(T_n^2(X,Y))=E(S_n^{2*}(X,Y))=Cov(X,Y)$$

$$E(S_n^2(X,Y))=\frac{n-1}{n}\operatorname{cov}(X,Y)$$

A-2 Exemple d'estimateurs

- **Estimateur de la corrélation ρ(X,Y) entre deux v.a. X et Y : La corrélation empirique**

$$\boxed{R_n(X,Y) = \frac{S_n(X,Y)}{S_n(X)S_n(Y)}}$$ est un estimateur de ρ(X,Y)

On note $r_n(X,Y)$ ou $\hat{\rho}(X,Y)$ l'estimation associée.

B- Précision d’un estimateur

B-1 Une mesure de précision : l'erreur quadratique moyenne (EQM)

Soit $\Theta_n \subset \mathbb{R}^p$ un estimateur de θ.

✓ Erreur Quadratique Moyenne : (EQM)

$$\boxed{E\left[(\Theta_n-\theta)^2\right]=V(\Theta_n)+(E(\Theta_n)-\theta)^2}$$

Variance=fluctuation aléatoire de Θ autour de sa moyenne

Biais^2=erreur systématique due au fait que Θ ne fluctue pas autour de θ

✓ Biais d'estimation :

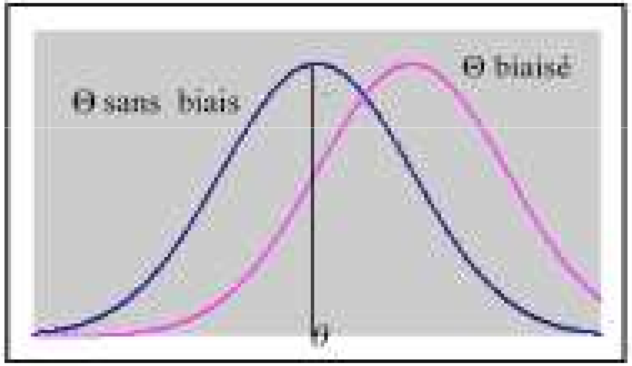
$$\boxed{B(\Theta_n)=E(\Theta_n)-\theta}$$

Un bon estimateur ponctuel doit être précis : il l'est d'autant plus que son erreur quadratique est faible : biais et variance les plus faibles possibles.

B-2 Absence de biais

✓ Un estimateur du paramètre θ est dit sans biais ssi :

$$\boxed{B(\Theta_n) = 0 \Leftrightarrow E(\Theta_n) = \theta}$$



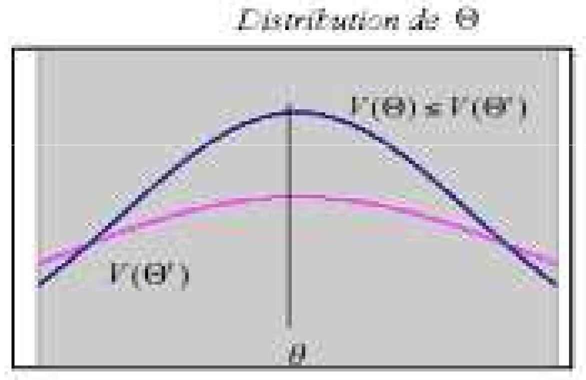
RQ: une propriété moins forte est l'absence de biais asymptotique

Ex : $\bar{X}_n$, T_n^2 et S_n^{*2} sont sans biais, S_n^2 est biaisé mais asymptotiquement sans biais

B-3 Variance minimale

- ✓ Un estimateur de θ est dit de variance minimale si, parmi tous les estimateurs possibles de θ il a la plus petite variance :

$$\boxed{\forall \Theta_n^* \text{ estimateur de } \theta,\ V(\Theta_n) \leq V(\Theta_n^*)}$$



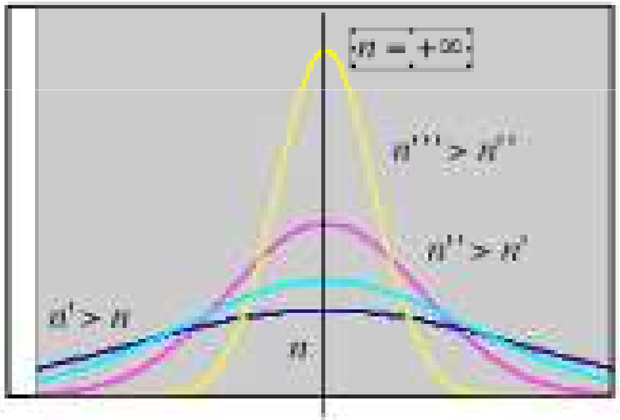
Pour deux estimateurs sans biais, le plus précis est celui qui a la plus petite variance.
Ex : lorsque E(X) =m est connue,

S_n^{*2} est moins précis que T_n^2

B-4 Convergence

- ✓ Convergence : un estimateur du paramètre θ est dit convergent ssi,

$$\boxed{\Theta_n \xrightarrow[n \to \infty]{P} \theta}$$



Pté : Un estimateur sans biais et de variance asymptotiquement nulle est convergent.

RQ : deux estimateurs convergents peuvent ne pas converger à la même vitesse

B-5 : Recherche du meilleur estimateur

- ✓ La propriété la plus désirable pour un estimateur est d'avoir une **faible Erreur Quadratique Moyenne** (ce qui n'exige pas forcément d'être sans biais).

Pbme : *la théorie de l'estimation ne permet pas de résoudre le problème de minimisation de l'EQM (fonction dépendant de manière complexe du paramètre)*

- ✓ **Un compromis : recherche d'un estimateur sans biais de variance minimale**

  - L'absence de biais facilite grandement l'étude des propriétés d'un estimateur car le biais d'un estimateur peut dépendre de façon complexe de la valeur du paramètre.

  - ***Rque : Il est cependant possible de trouver des estimateurs biaisés plus précis que le meilleur estimateur sans biais.***

*Pbme : le calcul de la variance d'un estimateur et donc l'existence et la définition d'un estimateur de variance minimale nécessite généralement la connaissance de la **loi de probabilité jointe de l'échantillon aléatoire**,*

B-5 : Recherche du meilleur estimateur

✓ **Loi de l'échantillon aléatoire dans un tirage aléatoire simple:**

X, $f_\theta(x)$ $\Rightarrow$ $(X_1,\ldots,X_n)$ i.i.d., $L_\theta(x_1,\ldots,x_n)$

$$\boxed{L_\theta(x_1,\ldots,x_n)=\prod_{i=1}^{n} f_\theta(x_i), \quad \forall(x_1,\ldots,x_n)\in \mathbb{R}^n}$$

- On connait généralement la forme paramétrique de f, mais elle dépend du paramètre q inconnu. Donc, la loi de l'échantillon est inconnue en général (dépend de paramètres inconnus que l'on cherche justement à estimer).

- Solution : Trouver un estimateur sans biais de variance minimale ne sera possible que si

➢ L'info contenue dans l'échantillon sur θ est suffisamment riche $\Rightarrow$ **information**

➢ On dispose d'une statistique exhaustive pour θ $\Rightarrow$ **exhaustivité**

B-5 : Recherche du meilleur estimateur

✓ Remarque s :

- L'absence de biais ne garantit pas la plus faible valeur possible de l'*EQM* : celle-ci sera atteinte lorsque sera trouvé le meilleur compromis entre le biais de l'estimateur et sa variance

- Dans certains cas même, l'introduction d'un léger biais dans un estimateur initialement sans biais peut conduire à une réduction significative de sa variance, au point de provoquer une diminution de son *EQM*, et donc d'améliorer ses performances.

C- Exhaustivité et information

C-1 Introduction

- ✓ **Information :** Un sondage (= une réalisation de l'échantillon aléatoire) nous apporte un certaine information sur θ (la répartition de ses valeurs nous donne une information sur la loi de X, qui dépend de θ). Elle doit être suffisante pour pouvoir espérer estimer θ.

- ✓ **Exhaustivité :** L'estimation de θ faite à partir de ce sondage perd forcement une partie de cette information : partant de n valeurs, on n'en construit qu'une seule, l'estimation. Et la connaissance de la seule estimation ne permet pas de remonter à l'échantillon tout entier.

$\Longrightarrow$ Affaiblissement de l'information sur la loi

- Via le sondage
- Via la construction d'un estimateur

La perte doit être minimale pour construire un estimateur précis

C-2 Vraisemblance d'un échantillon

✓ **Rappels de notations** :

- θ le vecteur de paramètre inconnu ($\theta \in R^p$)
- $f_\theta(x)$ la loi (continue ou discrète) de X
- $(X_1, \ldots, X_n)$ l'échantillon aléatoire associé au sondage
- $(x_1, \ldots, x_n)$ une réalisation de cet échantillon ((un sondage particulier)

✓ **Remarque** : La loi $f_\theta(x)$ de X (ainsi que toute fonction dépendant de θ) sera notée dans ce chapitre $f(x, \theta)$, car on s'intéressera aux variations de la loi par rapport à θ.

C-2 Vraisemblance d'un échantillon

✓ **Vraisemblance de l'échantillon aléatoire :**

$$\boxed{L(X_1,....,X_n,\theta)=\prod_{i=1}^{n} f(X_i,\theta)}$$

NB : C'est une variable aléatoire. Sa réalisation sur le jeu de données (x1, x2, ..., xn) , est la valeur de la densité de l'échantillon aléatoire au point (x1, x2, ..., xn) .

✓ **Log-vraisemblance de l'échantillon aléatoire :**

$$\boxed{l(X_1,....,X_n,\theta)=\ln(L(X_1,....,X_n,\theta))=\sum_{i=1}^{n}\ln(f(X_i,\theta))}$$

NB:

- Pour des résolutions mathématiques, il est plus commode d'étudier la log vraisemblance que la vraisemblance.
- L et l , considérées comme des fonctions de θ, ont le même sens de variation.

C-3 Information de Fisher d'un échantillon (θ réel)

✓ **Score de l'échantillon** : Si *f(x, θ) est* différentiable en θ, l est une fonction dérivable de θ, et le score est sa dérivée :

$$\boxed{S_n(\theta) = \frac{\partial}{\partial \theta} l(X_1,\ldots,X_n,\theta) = \frac{1}{L(X_1,\ldots,X_n,\theta)} \frac{\partial}{\partial \theta} L(X_1,\ldots,X_n,\theta)}$$

- Pour tout θ, Le score est une variable aléatoire

- Le score s'annule à un optimum en θ de la fonction de vraisemblance.

- Pour une réalisation donnée (x1,...,xn) de l'échantillon aléatoire (un sondage particulier) La valeur du score est une fonction de θ, réalisation de la variable aléatoire définie ci-dessus sur ce jeu de données

$$s_n(\theta) = \frac{\partial}{\partial \theta} l(x_1,\ldots,x_n,\theta)$$

C-3 Information de Fisher d'un échantillon (θ réel)

- Le score est centré : $E(S_n(\theta)) = 0$

- la variance du score (si elle existe) s'appelle l'information de Fisher apportée par l'échantillon sur θ :

$$\boxed{I_n(\theta) = E\left((S_n(\theta))^2\right)}$$

Cette quantité mesure l'information apportée par un échantillon sur le paramètre

C-3 Information de Fisher d'un échantillon (θ réel)

- **L'information de Fisher mesure l'information apportée par un échantillon sur le paramètre : une information de Fisher proche de zero indique un échantillon peu informatif sur la valeur de θ**

- Pour un sondage particulier, la valeur du score mesure la sensibilité de la vraisemblance à la valeur de θ. Si le score est faible, la vraisemblance est peu sensible à de petites variations du paramètre : les observations n'arrivent pas à s'accorder entre elles sur la direction du changement à apporter à la valeur de θ pour augmenter la vraisemblance de l'échantillon. On doit donc s'attendre à ce que l'échantillon contienne peu d'information sur la vraie valeur de ce paramètre.

- En moyenne pour un θ fixé, le score est nul. Si sa variance (information de Fisher) est très petite pour une valeur donnée de θ, alors, presque tous les jeux de données auront alors un score proche de 0 (l'espérance du score), et donc presque tous les échantillons ne contiendront qu'une faible quantité d'information sur la valeur réelle de θ .

C-3 Information de Fisher d'un échantillon (θ réel)

✓ **Propriétés de l'information de Fisher :**

➢ Autre formulation : si le domaine de définition de X ne dépend pas de θ et que cette quantité existe,

$$\boxed{I_n(\theta) = -E\left(\frac{\partial^2 l(X_1,\dots,X_n,\theta)}{\partial\theta^2}\right) = -E\left(\frac{\partial S_n(\theta)}{\partial\theta}\right)}$$

➢ Additivité : si le domaine de définition de X ne dépend pas de θ , chaque observation apporte la même information

$$\boxed{I_n(\theta) = nI_1(\theta)}$$

C-3 Information de Fisher d'un échantillon (cas multidimentionnel)

✓ **Extension au cas multidimensionnel :**

Le score est un vecteur aléatoire de dimension p: $\theta = (\theta_1, \ldots, \theta_p) \in R^p$

$$\boxed{S_n(\theta) = \left(\frac{\partial}{\partial \theta_1} l(X_1, \ldots, X_n, \theta), \ldots, \frac{\partial}{\partial \theta_p} l(X_1, \ldots, X_n, \theta) \right)}$$

Il est caractérisé par son vecteur espérance (=0) et sa matrice de variance covariance appelée matrice d'information de Fisher, $I_n(\theta) = (I_{i,j}^n)_{1 \le i, j \le p}$ définie positive de terme général

$$I_{i,j}^n = Cov\left(\frac{\partial l(X_1, \ldots, X_n, \theta)}{\partial \theta_i}, \frac{\partial l(X_1, \ldots, X_n, \theta)}{\partial \theta_j} \right)$$

C-4 Statistique exhaustive

- **Dégradation de l'information par une statistique de l'échantillon :** soit T une statistique de l'échantillon et g (t, θ) sa loi. Alors

$$\boxed{I_n(\theta) \geq I_T(\theta)}$$

avec égalité si la statistique est exhaustive.

Dém :

- La vraisemblance de T est $g(T,\theta)$
- l'information de Fisher $I_T(\theta)$ apportée par T sur θ est (sous de bonnes conditions) :

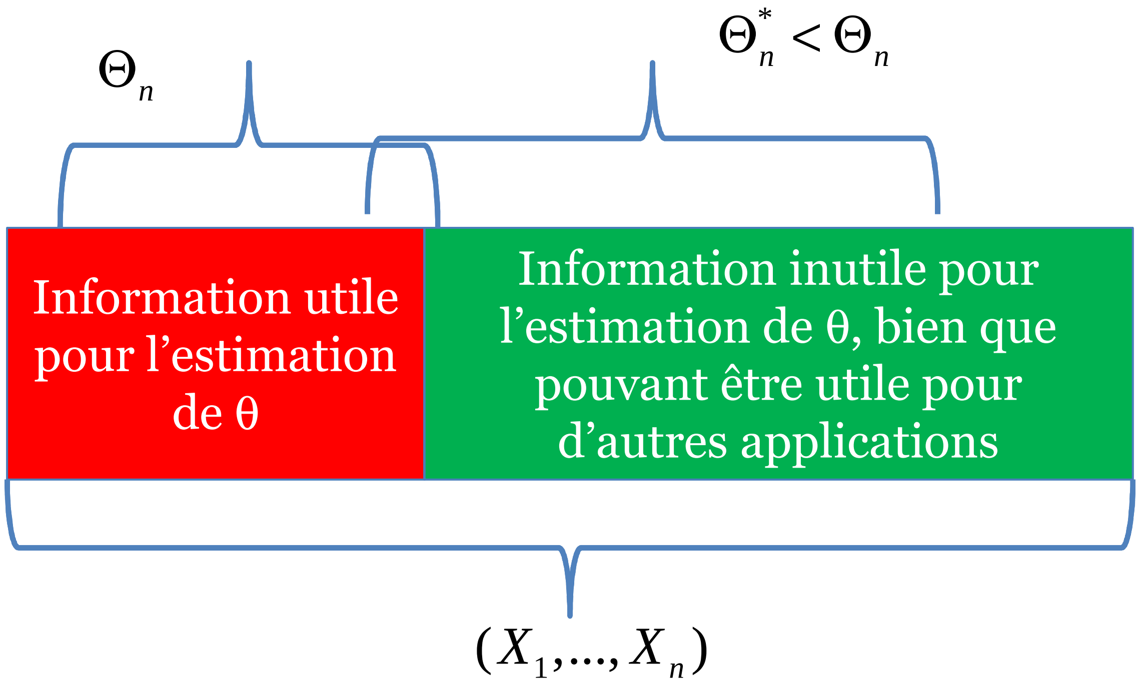
$$I_T(\theta) = -E\left(\frac{\partial^2 \ln g(T,\theta)}{\partial \theta^2}\right)$$

$$L(X_1,\ldots,X_n,\theta) = g(t,\theta)k(X_1,\ldots,X_n,\theta / T = t)$$ donc $$I_n(\theta) = I_T(\theta) - E\left(\frac{\partial^2 \ln k}{\partial \theta^2}\right)$$

Le dernier terme est l'information conditionnelle de l'échantillon sachant T et on montre qu'il est positif

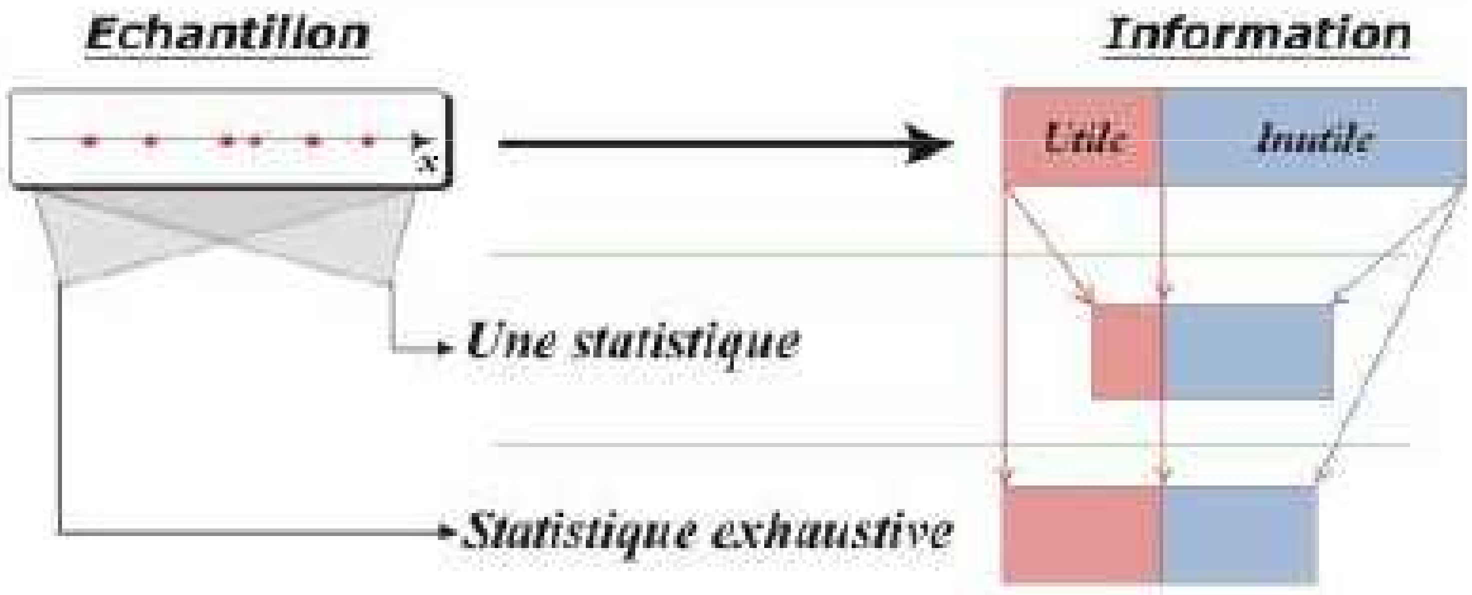
C-4 Statistique exhaustive

Une statistique de l'échantillon ne peut pas contenir plus d'information sur θ que l'échantillon ….Mais dans quel cas peut-on espérer qu'elle conserve la majeure partie de l'information « utile »?



C-4 Statistique exhaustive

✓ **Statistique exhaustive T**: sa création ne rejettera que de l'information "inutile" tout en préservant intégralement l'information "utile" à l'estimation de θ.



Construction de T?

C-4 Statistique exhaustive

✓ **Idées de construction** :

- Soit t la valeur de T sur un sondage $(x_1,\dots,x_n)$. Pour que t soit aussi informative que ce sondage, il faut à partir de la valeur t être capable de reconstruire $(x_1,\dots,x_n)$ ou une autre réalisation $(x'_1,\dots,x'_n)$ de $(X_1,\dots,X_n)$

Soit L_θ la loi de $(X_1,\dots,X_n)$. Soit g_θ la loi de T dont une réalisation est t. Pour tirer une réalisation de $(X_1,\dots,X_n)$ on peut :

- simuler une réalisation de loi L_θ
- simuler une réalisation t de loi g_θ, puis une réalisation de la loi k de l'échantillon conditionnellement à la valeur de t car :

$$L_\theta(x_1,\dots x_n) = k_\theta(x_1,\dots x_n / T = t)\, g_\theta(t)$$

C-4 Statistique exhaustive

Statisticien S1 :

- ✓ dispose d'un jeu de données obtenu par sondage.
- ✓ Peut construire la valeur de t à partir de ce jeu de données
- ✓ Peut construire une estimation de θ à partir de ce jeu de données.

Statisticien S2 :

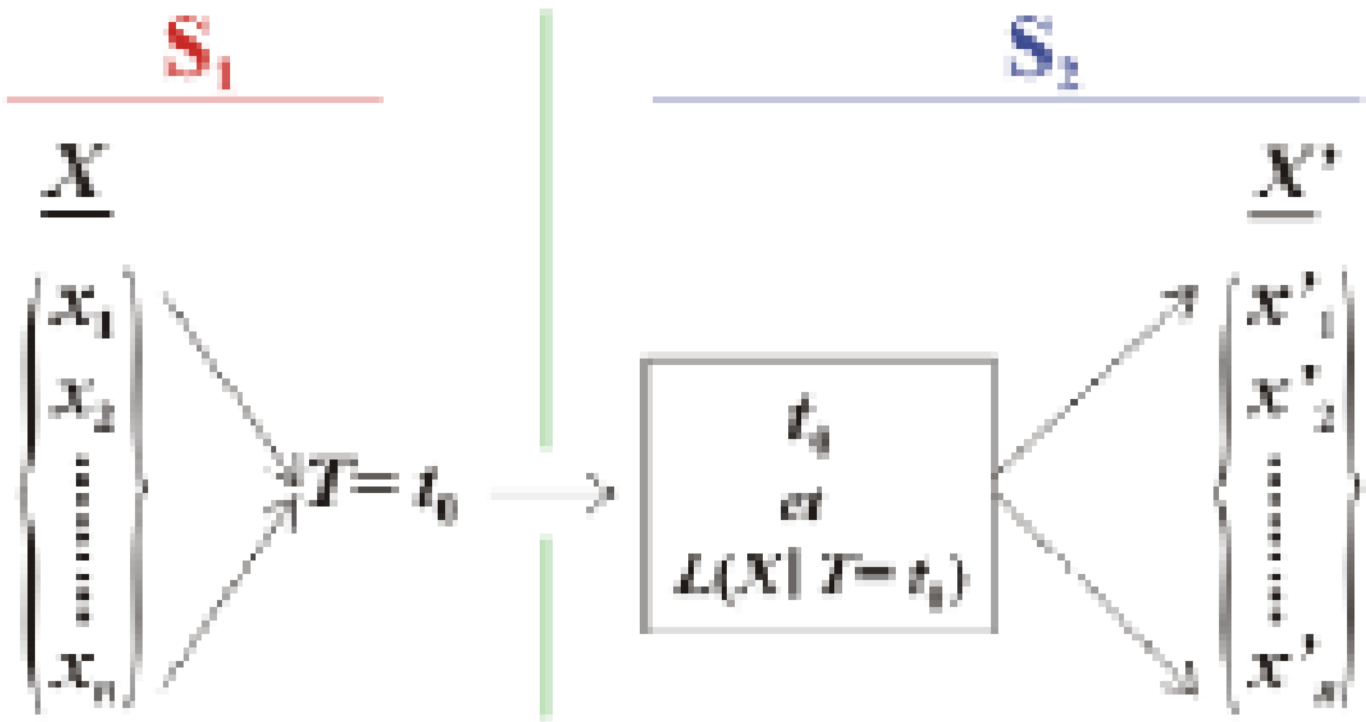
- ✓ ne dispose pas du jeu de données
- ✓ s'est fait donner t par S1
- ✓ Connait g_θ

Pour disposer d'autant d'information que S1, S2 doit être capable de tirer une réalisation de l'échantillon aléatoire. Comme il dispose de t, il faut qu'il puisse tirer dans la loi conditionnelle de cet échantillon sachant t, mais celle-ci dépend généralement de θ.

$$\boxed{\text{S1 et S2} \Leftrightarrow k_\theta(x_1,\ldots x_n / T = t) \text{ ne dépend pas de } \theta}$$

Alors seulement, S2 sera capable de tirer une réalisation de l'échantillon aléatoire et de construire une estimation de θ à partir de ce jeu de données.

C-4 Statistique exhaustive



X et X' ont des distributions identiques

C-4 Statistique exhaustive

- Une statistique T ne peut nous renseigner sur la valeur d'un paramètre que si sa loi dépend de ce paramètre. Puisque

$$L_\theta(x_1,\ldots x_n) = k_\theta(x_1,\ldots x_n / T = t)g_\theta(t)$$

Si la loi conditionnelle de l'échantillon aléatoire sachant la valeur de T ne dépend plus du paramètre, cela veut dire qu'alors, une fois T connu, nous n'obtenons plus aucune information sur le paramètre par l'échantillon et que donc T porte toute l'information disponible sur le paramètre.

C-4 Statistique exhaustive

✓ Formalisation

- Soit $L_\theta : \mathbb{R}^n \to \mathbb{R}$ (resp. $\mathbb{R}^n \to [0,1]$)

la densité jointe (resp. la loi de probabilité jointe) de l'échantillon aléatoire issu de X continue (resp.discrète).

- Soit $h : \mathbb{R}^n \to \mathbb{R}^p$ une fonction et $T = h(X_1,\dots,X_n)$ une statistique de l échantillon aléatoire et

  $t = h(x_1,\dots,x_n)$ sa valeur au point $(x_1,\dots,x_n)$

C-4 Statistique exhaustive

- **Définition :** T est dite exhaustive pour le paramètre θ si la distribution de l'échantillon aléatoire conditionnellement $T=t$ ne dépend pas de θ :

$$\boxed{L_\theta(x_1,\dots,x_n / T=t) = k(x_1,\dots,x_n / T=t), \quad \forall (x_1,\dots,x_n) \in \mathbb{R}^n, \forall \theta \in \vartheta}$$

- **Propriété :** Si T est exhaustive, alors $\boxed{I_n(\theta) = I_T(\theta)}$

- **Caractérisation (théorème de factorisation)**

Une statistique T de loi g_θ est dit **exhaustive** pour le paramètre θ ssi il existe une fonction k : $\mathbb{R}^n \to \mathbb{R}$ telle que :

$$\boxed{L_\theta(x_1,\dots,x_n) = g_\theta(t) h(x_1,\dots,x_n), \quad \forall (x_1,\dots,x_n) \in \mathbb{R}^n, \forall \theta \in \vartheta}$$

C-4 Statistique exhaustive

- **Exemple**

Si X suit une loi de poisson de paramètre θ, $T = \sum_{i=1}^{n} X_i$ *est une statistique exhaustive pour* θ. *En effet, T suit une loi de poisson de paramètre* $n\theta$, *donc*

$$g(t,\theta) = \frac{e^{-n\theta}(n\theta)^t}{t!} = \frac{e^{-n\theta}(n\theta)^{\sum_{i=1}^{n} x_i}}{(\sum_{i=1}^{n} x_i)!}$$

et

$$L(\mathbf{x},\theta) = \prod_{i=1}^{n} P_\theta(X = x_i) = \prod_{i=1}^{n} \frac{e^{-\theta}(\theta)^{x_i}}{x_i!} = \frac{e^{-n\theta}\theta^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$$

donc

$$h(\mathbf{x}) = \frac{L}{g} = \frac{(\sum_{i=1}^{n} x_i)!}{n^{\sum_{i=1}^{n} x_i}\prod_{i=1}^{n} x_i!} = \frac{t!}{n^t\prod_{i=1}^{n} x_i!} \quad (\textit{ne dépend pas de } \theta)$$

C-4 Statistique exhaustive

- **Lois de X permettant une statistique exhaustive (théorème de Darmois)**

On note f_θ la loi de X (densité si X continue, loi de probabilité si X discrète). On suppose que le domaine de définition de X ne dépend pas de θ.

a, b, α,β étant des fonctions,

- une CNS pour que l'échantillon aléatoire admette une statistique exhaustive pour θ est que

$$\boxed{f_\theta(x) = \exp\left[a(x)\alpha(\theta) + b(x) + \beta(\theta)\right]}$$

(famille exponentielle)

- Dans ce cas, si l'application $x_1 \to \sum_{i=1}^{n} a(x_i)$ est bijective et continûment différentiable pour tout xi , alors $T = \sum_{i=1}^{n} a(X_i)$ est une statistique exhaustive particulière pour θ.

C-4 Statistique exhaustive

- ✓ **Conséquences de l'existence d'une statistique exhaustive de θ sur l'existence et la définition d'un estimateur de θ de variance minimale :**

  Un estimateur de θ devrait être d'autant plus précis qu'il capture, pour sa construction, une part importante de l'info sur θ contenue dans l'échantillon. Dans la meilleure des situation, On devrait pouvoir retrouver toute l'info de l'échantillon à partir de l'estimateur

  $\Rightarrow$ les estimateurs les plus précis de θ (en particulier les estimateurs sans biais de variance minimale) sont des statistiques exhaustives ou des fonctions de celles-ci., si tant est qu'une statistique exhaustive de θ existe.

- ✓ **Un autre aspect du problème : l'information**

  Ce qui précède n'est pas suffisant pour avoir un estimateur précis : un estimateur ne pourra être bon que si l'échantillon lui-même sur lequel il est construit véhicule (contienne) suffisamment d'info sur le paramètre

  $\Rightarrow$ il faut en plus que l'**information contenue dans l'échantillon sur le paramètre soit suffisante**.

D- Estimateur de variance minimale, estimateur efficace

D-1 Estimateur sans biais de Variance Minimale

Rappel : Il est fréquent qu'un paramètre admette plusieurs, voire une infinité d'estimateurs sans biais.

Ex : Pour la distribution normale $N(\mu, s)$, la moyenne et la médiane empiriques sont toutes deux des estimateurs sans biais de μ) .

- De tous les estimateurs sans biais de θ, le meilleur (au sens de EQM) est celui qui a la plus faible variance. On l'appelle "Estimateur sans biais de Variance Minimale" de θ.

- L'identification et la qualité d'un estimateur sans biais de variance minimale est lié à l'information contenue dans l'échantillon sur θ et à l'existence d'une statistique exhaustive pour θ.

On dispose de quatre résultats théoriques pour identifier cet estimateur.

D-1 Estimateur sans biais de Variance Minimale

- **Unicité : s'il existe un estimateur sans biais de variance minimale de θ, alors, il est unique p.s.**

- **Théorème de Rao-Blackwell**

  Si un estimateur sans biais de θ n'est pas de variance minimale, il est possible de l'améliorer si l'on dispose d'une statistique exhaustive pour θ.

  Le Théorème ne garantit cependant pas que le nouvel estimateur "amélioré" soit de variance minimale.

- **Inégalité de Cramér-Rao**

  Permet d'établir, sous condition de régularité, une borne inférieure de la variance d'un estimateur sans biais.

- **Conditions sous lesquelles la borne est atteinte**. L'estimateur de variance minimale est alors celui ayant la variance de la borne de Cramer-Rao, il s'appelle estimateur efficace.

D-2 Théorème de Rao-Blackwell

- ✓ **Problème posé** : Si on dispose d'un estimateur sans biais de θ, est-il possible de l'améliorer?
- ✓ **Réponse** : oui si l'on dispose d'une statistique exhaustive de θ,

❑ **Théorème de Rao-Blackwell** : Soit Θ_n un estimateur sans biais de θ et T une statistique exhaustive de θ. Alors $\Theta_n{}^*=E(\Theta_n/T)$ est un estimateur sans biais de θ au moins aussi bon que Θ_n.

❑ **Conséquence** : s'il existe une statistique exhaustive T de θ alors l'estimateur sans biais de variance minimale de θ est une fonction de T: $\Theta_n{}^*=k(T)$.

Rq : Inversement si on dispose d'un estimateur sans biais fonction d'une statistique exhaustive, on n 'est pas sûrs qu'il soit de variance minimale.

D-2 Théorème de Rao-Blackwell

- **Idée de la preuve**: on utilise les résultats suivants avec $Y = \Theta_n$ et Z=T et Θ_n^{**} un estimateur sans biais quelconque de θ,

$$\boxed{\begin{array}{l} \forall Y, Z \quad v.a. \\ E(E(Y \mid Z)) = E(Y) \\ \text{Théorème de la variance totale :} \\ V(Y) = V(E(Y \mid Z)) + E(V(Y \mid Z) \end{array}}$$

Pour la conséquence : Soit $\Theta_n^{**} = E(\Theta_n^* \mid T))$. Comme Θ_n^* est de variance minimale, on a $E(V(\Theta_n^* \mid T)) = 0$ et donc $\Theta_n^* = k(T)$

D-3 Inégalité de FDCR

- ✓ **Problème posé** : Etant donné un paramètre θ d'une distribution, quelle est la plus petite variance que l'on puisse espérer pour un estimateur sans biais de θ (ou d'une fonction k(θ) de θ ?
- ✓ **Réponse :** cela dépend de l'information contenue dans l'échantillon

H : Le domaine de définition de X ne dépend pas de θ et l'information de Fisher existe

❑ **Inégalité de Fréchet-Darmois-Cramer-Rao (FDCR):**

Si H est satisfaite, on a pour tout estimateur sans biais Θ_n de θ :

$$\boxed{V(\Theta_n) \geq \frac{1}{I_n(\theta)}}$$

.

D-3 Inégalité de FDCR

Interprétation : **La valeur de la borne de FDCR est fonction de l'information que peut contenir l'échantillon sur le paramètre** : Plus grande est l'information sur la valeur du paramètre, plus précises seront les prédictions d'un estimateur sans biais dont la variance est égale à la borne de Cramér-Rao . Inversement, si la variance du score (information de Fisher) est très petite, et donc si presque tous les échantillons ne contiennent que peu d'information sur la valeur du paramètre , on ne peut pas espérer d'un estimateur sans biais qu'il soit précis, c'est à dire qu'il ait une faible variance.

Remarque : L'inégalité de FDCR et donc la variance minimale que peut atteindre un estimateur sans biais, n'est valable que dans le cas où l'information de Fisher existe et ou H est vérifiée.

D-3 Inégalité de FDCR

- **Généralisation à un estimateur sans biais d'une fonction de θ:**

  soit k une fonction et Δ_n un estimateur sans biais de k(θ). Si k est une fonction dérivable et que H est satisfaite:

$$\boxed{V(\Delta_n) \geq \frac{\left(k'(\theta)\right)^2}{I_n(\theta)}}$$

D-4 Estimateur efficace

✓ **Estimateur efficace**

- **Définition** : un estimateur efficace de θ (resp. k(θ)) est un estimateur sans biais dont la variance est égale à la borne inférieure de Cramer-Rao :

$$\boxed{V(\Theta_n) = \frac{1}{I_n(\theta)}} \qquad \left(\text{resp. } \boxed{V(\Delta_n) = \frac{\left(k'(\theta)\right)^2}{I_n(\theta)}}\right)$$

- **Propriétés** :

➢ Si un estimateur efficace de θ (ou de k(θ)) existe, il est unique p.s.

➢ Si un estimateur efficace de θ (resp. k(θ)) existe, il est égal p.s. à l'estimateur sans biais de variance minimale de θ (resp. k(θ).

- **Pertinence** : La définition d'un estimateur efficace n'a de sens que si l'hypothèse H est vérifiée.

D-4 Estimateur efficace

Exemple : Loi de Poisson de paramètre θ :

$$\boxed{\begin{aligned}
&l_n(X_1,\ldots,X_n,\theta) = -n\theta + n\overline{X}\ln\theta - \ln\left(\prod_{i=1}^{n} X_i!\right)\\
&S_n(\theta) = -n + \frac{n\overline{X}}{\theta}\\
&I_n(\theta) = -E\left(\frac{\partial}{\partial\theta}S_n(\theta)\right) = \frac{nE(\overline{X})}{\theta^2} = \frac{n}{\theta}
\end{aligned}}$$

- ✓ On estime θ par $\overline{X}$: on a $V(\overline{X}) = \dfrac{\theta}{n}$. C'est un estimateur efficace.

D-4 Estimateur efficace

- **<u>Problème posé</u>** : Conditions nécessaire à l'existence d'un estimateur efficace de $k(\theta)$

- **<u>Réponses</u> :** La loi de X permet une statistique exhaustive (Darmois) de θ . Dans ce cas, l'estimateur efficace de $k(\theta)$ est une statistique exhaustive de θ

Dans la suite, on suppose que H est vérifiée.

D-4 Estimateur efficace

- **Théorème sur l'efficacité**

- La borne de Cramer-Rao ne peut être atteinte (il ne peut exister d'estimateur efficace) que si la loi de X appartient à la famille exponentielle

$$\boxed{f_\theta(x) = \exp\left[a(x)\alpha(\theta) + b(x) + \beta(\theta)\right]}$$

(c'est-à-dire si elle permet l'existence d'une statistique exhaustive, Car l'estimateur efficace est nécessairement exhaustif pour θ.)

- Si la loi de X est bien de la forme précédente, il n'existe qu'une seule fonction de θ qui puisse être estimée efficacement, c'est : $\boxed{k(\theta) = \frac{-\beta'(\theta)}{\alpha'(\theta)}}$

- L'estimateur de k(θ) est alors $\boxed{T_n = \frac{1}{n}\sum_{i=1}^{n} a(X_i)}$ de variance (minimale) :

$$V(T_n) = \frac{(k'(\theta))^2}{I_n(\theta)} = \frac{k'(\theta)}{n\alpha'(\theta)}$$

D-4 Estimateur efficace

Exemple : Loi de Poisson de paramètre θ :

On a vu que : $S = \sum_{i=1}^{n} X_i$ est exhaustive.

On peut le retrouver autrement : $\ln f(x,\theta) = -\theta + x\ln\theta - \ln(x!)$ donc la loi de poisson appartient à la famille exponentielle avec

$$a(x) = x; \alpha(\theta) = \ln\theta; b(x) = -\ln(x!); \beta(\theta) = -\theta$$

Donc d'après le théorème de Darmois $S = \sum_{i=1}^{n} a(X_i) = \sum_{i=1}^{n} X_i$ est exhaustive.

D'après le théorème précédent la seule fonction qui puisse être estimée efficacement est $\boxed{k(\theta) = \dfrac{-\beta'(\theta)}{\alpha'(\theta)} = \theta}$ l'estimateur efficace est S/n

Remarque :Si S est exhaustive, toute fonction déterministe de S est exhaustive. Par exemple, S/n est exhaustive.

D-4 Estimateur efficace

- **Remarque importante**: Il existe au plus une seule fonction k(θ) du paramètre θ qui peut être estimée efficacement. En conséquence, s'il existe une fonction k vérifiant la relation ci-dessus, et si cette fonction n'est **pas la fonction identité, alors il n'existe pas d'estimateur efficace de θ.**

D-4 Estimateur efficace

- **Conclusion sur l'estimation de θ :**

Cas où les hypothèses de L'inégalité de Cramér-Rao sont satisfaites :

L'inégalité de Cramér-Rao produit une borne inférieure de la variance d'un estimateur sans biais. Un estimateur sans biais de variance minimale aura donc une variance supérieure ou égale à la borne inférieure de l'inégalité de FDCR.

Cette borne n'est pas forcément atteinte dans le cas général : On peut par exemple exhiber des cas où cette borne est égale à 0, et donc évidemment inaccessible.

Lorsqu'elle l'est, On appelle estimateur efficace un estimateur sans biais dont la variance est égale à la borne inférieure de l'inégalité de FDCR. Dans ce cas, c'est l'estimateur de variance minimale. Il n'existe que si X admet une loi dans la famille exponentielle et que k est la fonction indentité, et l'estimateur est dans ce cas unique et exhaustif.

Cas où les hypothèses de L'inégalité de Cramér-Rao ne sont sas satisfaites :

La notion d'estimateur efficace n'a alors pas de sens, ni la borne de Cramer-Rao. Il peut alors même exister des estimateurs sans biais dont les variances sont inférieures à la borne (sans signification) de Cramér-Rao.

E- Quelques méthodes d’estimation

E-1 Estimation par la méthode du maximum de vraisemblance

Soit **(** $X_1, X_2, ..., X_n$) l'échantillon aléatoire associé à un sondage de taille n dans la population de distribution de probabilité $f(x, \theta)$, où θ est un vecteur de paramètres inconnus, $\theta \in \vartheta \subset R^p$ et que l'on cherche à estimer.

- ✓ **Estimation par la méthode du maximum de vraisemblance (MLE)** : On choisit comme estimateur de θ le vecteur θ_n qui rend maximal la vraisemblance (ou la log-vraisemblance) de l'échantillon dont on dispose.

$$\boxed{\Theta_n = \arg\max_{\theta \in \vartheta} L(X_1, ..., X_n, \theta) = \arg\max_{\theta \in \vartheta} l(X_1, ..., X_n, \theta)}$$

Pour un sondage particulier, **(** $x_1, x_2, ..., x_n$), l'estimation (valeur) du paramètre est :

$$\boxed{\hat{\theta}_n = \arg\max_{\theta \in \vartheta} L(x_1, ..., x_n, \theta) = \arg\max_{\theta \in \vartheta} l(x_1, ..., x_n, \theta)}$$

E-1 Estimation par la méthode du maximum de vraisemblance

✓ **Mise en œuvre**

Lorsque $f(x, \theta)$ est une fonction deux fois différentiable en θ, la méthode la plus directe consiste à :

- Identifier les extrema de *la vraisemblance (ou log-vraisemblance)* en annulant ses dérivées partielles premières par rapport à θ (le score) . On résout donc en θ le système d'équations : $\boxed{S_n(\theta) = 0}$
- Retenir parmi ces extrema ceux qui sont des maxima, par exemple en recherchant ceux pour lesquels la matrice des dérivées partielles secondes de *la vraisemblance (ou log-vraisemblance)* est négative au voisinage de θ.
- Retenir, de ces différents maxima, celui qui présente la plus grande valeur de la vraisemblance.

E-1 Estimation par la méthode du maximum de vraisemblance

- ✓ **Exemple : EMV de l'espérance d'une loi exponentielle de paramètre** θ

$$L_n(X_1,\dots,X_n,\theta)=\prod_{i=1}^{n}\theta e^{-\theta X_i}=\theta^n e^{-\theta n\bar{X}}$$

$$l_n(X_1,\dots,X_n,\theta)=n\ln\theta-\theta n\bar{X}$$

$$S_n(\theta)=\frac{n}{\theta}-n\bar{X}\Rightarrow\Theta_n=\frac{1}{\bar{X}}$$

$$\frac{\partial}{\partial\theta}S_n(\theta)=-\frac{n}{\theta^2}<0$$

E-1 Estimation par la méthode du maximum de vraisemblance

✓ **Propriétés de l'estimateur du Maximum de Vraisemblance (EMV) (à titre indicatif)**

❑ **Propriétés à taille d'échantillon fixe**

Prop 1 : S'il existe un estimateur efficace de θ , alors il est égal à l'unique EMV de θ.

RQ : La réciproque est fausse : un EMV n'est pas obligatoirement efficace.

Prop 2 : Si le paramètre θ admet une statistique exhaustive, l'EMV est une fonction de cette statistique exhaustive.

Prop 3 : Si Θ est l'EMV de θ, $k(\Theta)$ est l'EMV de $k(\theta)$.

E-1 Estimation par la méthode du maximum de vraisemblance

- **Propriétés asymptotiques**

Prop 4 : Pour des échantillons suffisamment grands, l'EMV devient unique, et tend (en probabilité) vers la vraie valeur du paramètre θ. C'est donc un estimateur convergent :

$$\hat{\theta}_n \xrightarrow{P} \theta$$

Prop 5 : L'EMV est asymptotiquement gaussien et asymptotiquement efficace (pour des échantillons suffisamment grands, sa variance est inférieure à celle de tout autre estimateur et est proche de la borne de Cramér-Rao) :

$$\frac{\hat{\theta}_n - \theta}{\frac{1}{\sqrt{I_n(\theta)}}} \xrightarrow{L} N(0,1)$$

où I_n désigne l'information de Fisher de l'échantillon.

E-1 Estimation par la méthode du maximum de vraisemblance

✓ Extensions : Vraisemblance et tests

Puisque la vraisemblance mesure la plus ou moins bonne adéquation entre une distribution de paramètre donné θ et un échantillon, on peut s'attendre à la voir jouer un rôle important dans les tests portant sur le choix entre des distributions paramétriques candidates pour rendre compte d'un échantillon. L'exemple le plus simple d'une telle apparition de la vraisemblance dans le monde des tests est le Théorème de Neyman-Pearson (cf chapitre suivant) qui établit que la Meilleure Région Critique d'un test devant décider entre deux distributions candidates est entièrement déterminée par des considérations portant sur les vraisemblances de ces deux distributions pour l'échantillon disponible.

E-2 Estimation par la méthode des moments

Soit ($X_1, X_2, \ldots, X_n$) l'échantillon aléatoire associé à un sondage de taille n dans la population de distribution de probabilité $f(x, \theta)$, où θ est un vecteur de paramètres inconnus de dimension p et que l'on cherche à estimer.

Soient

$$\boxed{\overline{X^r} = \frac{1}{n}\sum_{i=1}^{n} X_i^r \text{ et } m_r(\theta) = E_\theta(X^r)}$$

les moments empiriques et théoriques d'ordre r de l'échantillon et de la loi de X respectivement (lorsque r=1, on a la moyenne empirique et l'espérance). Les moments théoriques dépendent de θ.

.

E-2 Estimation par la méthode des moments

- ✓ **Estimateur des moments de θ :**

  L'estimateur des moments de θ est le vecteur Θ_n qui vérifie le système d'équations en θ:

$$\boxed{\overline{X^r} = m_r(\theta), \quad r = 1, \ldots p}$$

Si ce système peut être résolu,

$$\boxed{\Theta_n = k\left(\overline{X^1}, \ldots, \overline{X^p}\right)}$$

est l'estimateur des moments de θ.

Pour un sondage particulier **(** $x_1, x_2, \ldots, x_n$), l'estimation (valeur) correspondante est :

$$\boxed{\hat{\theta}_n = k\left(\overline{x^1}, \ldots, \overline{x^p}\right)}$$

E-2 Estimation par la méthode des moments

- ✓ **Exemple : EM de l'espérance d'une loi exponentielle de paramètre** θ

Θ_n solution de :

$$E_\theta(X) = \bar{X} \Leftrightarrow \frac{1}{\theta} = \bar{X} \Rightarrow \Theta_n = \frac{1}{\bar{X}}$$

E-2 Estimation par la méthode des moments

✓ **Propriétés (à titre indicatif)**

- La méthode des moments fournit des estimateurs convergents
- La méthode des moments est conceptuellement plus simple que la méthode du maximum de vraisemblance

- La méthode des moments fournit des estimateurs peu précis lorsque n est modéré
- les estimateurs ainsi produits n'ont pas les bonnes propriétés asymptotiques des estimateurs du Maximum de Vraisemblance. En particulier on n'a pas en général la loi limite de l'estimateur