EISTI

Ecole Internationale des Sciences du Traitement de l'Information

CHAPITRE 2 : LA VRAISEMBLANCE

Table des matières

Soit un échantillon X_1, ..., X_n sont des variables aléatoires i.i.d.. On suppose de plus que la loi de l'échantillon $f(x_1,\ldots,x_n\,;\theta)$ dépend d'un paramètre θ (vecteur ou scalaire).

1. DEFINITION

On définit la *fonction de vraisemblance* de l'échantillon pour le paramètre θ,

$$\theta \mapsto L(x_1,\ldots,x_n;\theta)$$

par la loi conjointe du vecteur $(X_1,\ldots,X_n)$.

Etant donné que la variables aléatoires sont indépendantes, cela revient à faire le produit des lois marginales, d'où, dans le cas discret,

$$L(x_1,\ldots,x_n;\theta)=\prod_{i=1}^{n} p_\theta(x_i)\,,$$

avec p_θ fonction de masse des X_i, et dans le cas continu,

$$L(x_1,\ldots,x_n;\theta)=\prod_{i=1}^{n} f_\theta(x_i)\,,$$

avec f_θ fonction de densité des X_i.

<u>Exemple</u>
Supposons que l'échantillon suit un loi de Poisson de paramètre θ, alors

$$L(x_1,\ldots,x_n;\theta)=\prod_{i=1}^{n}\frac{e^{-\theta}\theta^{x_i}}{x_i!}=e^{-n\theta}(\theta^{\sum_{i=1}^{n}x_i})(\prod_{i=1}^{n}\frac{1}{x_i!}).$$

Supposons que l'échantillon suit un loi exponentielle de paramètre θ, alors

$$L(x_1,\ldots,x_n;\theta)=\prod_{i=1}^{n}\theta e^{-\theta x_i}=\theta^n(e^{-\theta\sum_{i=1}^{n}x_i}).$$

2. INFORMATION DE FISHER

2.1. Définition

La fonction de vraisemblance est une fonction de θ, on peut donc la dériver et ainsi définir la *quantité d'information de Fisher* sur θ apporté par l'échantillon $X_1,\ldots,X_n$,

$$I_n(\theta)=E\left[\left(\frac{\partial}{\partial\theta}\ln L(X_1,\ldots,X_n;\theta)\right)^2\right],$$

si cette quantité existe.

On peut montrer que dans le cas où le support des X_i ne dépend pas de θ,

$$I_n(\theta)=-E\left[\frac{\partial^2}{\partial\theta^2}\ln L(X_1,\ldots,X_n;\theta)\right].$$

Cette définition est générale et dans le cas de l'échantillonnage, *i.e* de variables aléatoires i.i.d, il est facile de voir que

$$I_n(\theta)=nI_1(\theta).$$

<u>Exemple</u>
Reprenons le cas où l'échantillon suit une loi exponentielle de paramètre θ.

$$\ln L(x_1,\ldots,x_n;\theta)=n\ln\theta-\theta\sum_{i=1}^{n}x_i \;\Rightarrow\; \frac{\partial}{\partial\theta}\ln L(x_1,\ldots,x_n;\theta)=\frac{n}{\theta}-\sum_{i=1}^{n}x_i$$

$$\Rightarrow\; \frac{\partial^2}{\partial\theta^2}\ln L(x_1,\ldots,x_n;\theta)=-\frac{n}{\theta^2}$$

$$\Rightarrow\; I_n(\theta)=-E\left[-\frac{n}{\theta^2}\right]=\frac{n}{\theta^2}$$

2.2. Inégalité de Fréchet-Darmois-Cramer-Rao (FDRC)

La quantité d'information de Fisher fournit une borne minimale pour tout estimateur de θ par le théorème suivant.

Soit T_n un estimateur de $h(\theta)$ avec h dérivable, alors dans le cas où le support des X_i ne dépend pas de θ,

$$\operatorname{var}(T_n)\geq\frac{(h'(\theta))^2}{I_n(\theta)}.$$

Le terme de droite est appelé la *borne de Cramer-Rao*.

Ainsi tout estimateur aura une variance supérieure à la borne de Cramer-Rao. On peut noter que si la borne de Cramer-Rao est grande, il est impossible d'estimer θ correctement.

2.3. Efficacité d'un estimateur

Grâce au risque quadratique, il est possible de comparer des estimateurs entre eux. Se pose maintenant la question de savoir si l'estimateur utilisé est le meilleur.

On dit qu'un estimateur est *efficace* si sa variance est égale à la borne de Cramer-Rao.

Propriété
Un estimateur sans biais efficace est le meilleur estimateur de θ.

En effet, il est de variance minimale donc meilleur que les autres. De plus il est unique. Malheureusement, il n'existe pas toujours.

Exemple
Reprenons le cas où l'échantillon suit une loi exponentielle de paramètre θ et soit T_n un estimateur quelconque de θ, alors

$$\mathrm{var}(T_n) \geq \frac{1}{I_n(\theta)} = \frac{\theta^2}{n}.$$

Considérons maintenant $T_n = \overline{X}$ un estimateur de $1/\theta$ car $E(T_n)=E(X_i)=1/\theta$.
Alors

$$\frac{(h'(\theta))^2}{I_n(\theta)} = \frac{1}{\theta^4} \times \frac{\theta^2}{n} = \frac{1}{n\theta^2} \quad \text{et} \quad \mathrm{var}(\overline{X}) = \frac{\mathrm{var}(X_i)}{n} = \frac{1}{n\theta^2}.$$

On a donc $\mathrm{var}(T_n) = \frac{1}{I_n(\theta)}$ ($\theta \neq 1$) donc l'estimateur est efficace.

Cas des estimateurs usuels [1]

- Etant donné un échantillon de loi de Bernoulli B(p), alors la fréquence est un estimateur efficace de p.
- Etant donné un échantillon de loi normale $N(\mu,\sigma^2)$, alors la moyenne est un estimateur efficace de μ et la variance empirique n'est pas un estimateur efficace de σ^2.

3. ESTIMATEUR DU MAXIMUM DE VRAISEMBLANCE

Dans le cas où aucun des estimateurs usuels ne peut être *a priori* utilisé, on fait appel à un *estimateur du maximum de vraisemblance*. Comme son nom l'indique,

[1] Démonstration en exercice

cet estimateur maximise le logarithme de la fonction de vraisemblance. Il est donc défini comme une solution de l'équation,

$$\frac{\partial}{\partial\theta}\ln L(x_1,\ldots,x_n;\theta)=0\,.$$

Les propriétés (biais, convergences,...) de l'estimateur du maximum de vraisemblance sont inconnues et il faut donc les étudier à chaque cas. Seule la loi asymptotique est connue,

$$T_n \underset{n\to+\infty}{\overset{\mathcal{L}}{\longrightarrow}} N(\theta(\frac{1}{I_n(\theta)})$$

où T_n est l'estimateur du maximum de vraisemblance du paramètre θ. On peut donc en conclure que T_n est asymptotiquement sans biais et asymptotiquement efficace.

<u>Propriété</u>
Notons T_n l'estimateur du maximum de vraisemblance d'un paramètre θ. Alors $g(T_n)$ est un estimateur de $g(\theta)$.

<u>Exemple</u>
Reprenons le cas d'un échantillon de loi de Poisson de paramètre θ. Nous avons vu que

$$\ln L(x_1,\ldots,x_n;\theta)=-n\theta+\ln(\theta)\sum_{i=1}^{n}x_i+\ln(\prod_{i=1}^{n}\frac{1}{x_i!}) \;\Rightarrow\; \frac{\partial}{\partial\theta}\ln L(x_1,\ldots,x_n;\theta)=-n+\frac{1}{\theta}\sum_{i=1}^{n}x_i$$

D'où

$$\frac{\partial}{\partial\theta}\ln L(x_1,\ldots,x_n;\theta)\geq 0 \;\Leftrightarrow\; \theta\leq\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}$$

Ainsi $\bar{x}$ est un maximum et $\bar{X}$ est donc l'estimateur du maximum de vraisemblance de θ. On remarque que cet estimateur est sans biais puisque $E(\bar{X})=E(X_i)=\theta$. Le risque quadratique, $R_\theta=\mathrm{var}(\bar{X})=\theta/n$ tend vers 0 donc l'estimateur est convergent. De plus, il est efficace, en effet,

$$\frac{\partial^2}{\partial\theta^2}\ln L(x_1,\ldots,x_n;\theta)=-\frac{1}{\theta^2}\sum_{i=1}^{n}x_i \;\Rightarrow\; I_n(\theta)=-E[\frac{\partial^2}{\partial\theta^2}\ln L(X_1,\ldots,X_n;\theta)=\frac{1}{\theta^2}\sum_{i=1}^{n}E[X_i]=\frac{n}{\theta}$$

d'où

$$\mathrm{var}(\bar{X})=\frac{1}{I_n(\theta)}\,.$$

<u>Cas particuliers des estimateurs usuels</u> [1]

- dans le cas d'un échantillon de loi de Bernoulli $B(p)$, alors la fréquence empirique est l'estimateur du maximum de vraisemblance de p.
- dans le cas d'un échantillon gaussien $N(\mu,\sigma^2)$, alors la moyenne et la variance empirique sont les estimateurs du maximum de vraisemblance de μ et σ^2 respectivement.

4. METHODE DE NEYMAN-PEARSON DANS LES TESTS D'HYPTOTHESES

Il n'est pas toujours évident de deviner graphiquement l'allure de la région critique (par exemple dans le cas d'une loi du χ^2). La méthode de Neyman-Pearson permet de déterminer la meilleure variable de décision et la région critique optimale d'un test. Il s'agit en fait de maximiser la puissance 1-β pour une valeur donnée de α.

Supposons que la fonction de densité de l'échantillon est une fonction f(x,θ) où θ est le paramètre à tester et notons $L(x_1,\ldots,x_n\,;\theta)$ la fonction de vraisemblance de l'échantillon. Soit W une région critique de R^n. On a alors

$$\alpha = P(W|H_0) = \int_W L(x_1,\ldots,x_n;\theta_0)dx_1\ldots dx_n$$

fixé et on maximise

$$1-\beta = P(W|H_1) = \int_W L(x_1,\ldots,x_n;\theta_1)dx_1\ldots dx_n\ .$$

Ce problème a été résolu par Neyman et Pearson et on a le théorème suivant

<u>Théorème de Neyman-Pearson</u>

La région critique optimale est définie par l'ensemble des points de R^n tels que

$$\frac{L(x_1,\ldots,x_n;\theta_1)}{L(x_1,\ldots,x_n;\theta_0)} > k_\alpha\ .$$

<u>Exemple</u>

Lors d'une étude sur la fiabilité d'une certaine machine, on s'intéresse au nombre de ses défaillances (noté X) dans une journée. On sait que X suit une loi de Poisson :

$$P(X=x) = \frac{e^{-m}m^x}{x!},\ x=0,1,\ldots.$$

où m est un paramètre inconnu. Après 2 mois de relevés quotidiens, on observe que $\bar{x}=3.1$. On souhaite alors tester les hypothèses :

$$\begin{cases} H_0 : m=m_0=2 & \text{(fonctionnement normal)} \\ H_1 : m=m_1=5 & \text{(fonctionnement défaillant)} \end{cases}$$

en considérant un risque de $1^{ère}$ espèce de 5%.

La fonction de vraisemblance est définie par

$$L(x_1,\ldots,x_n;m) = e^{-nm}m^{\sum_{i=1}^n x_i} \times \left(\prod_{i=1}^n x_i!\right)^{-1}.$$

Le rapport des fonctions de vraisemblance sous les hypothèses H_0 et H_1 est égale

$$\frac{L(x_1,\ldots,x_n;\theta_1)}{L(x_1,\ldots,x_n;\theta_0)} = e^{-n(m_1-m_0)}\left(\frac{m_1}{m_0}\right)^{\sum_{i=1}^n x_i} > k_\alpha\ .$$

En passant au logarithme on a

$$\ln\left(\frac{L(x_1,\ldots,x_n;\theta_1)}{L(x_1,\ldots,x_n;\theta_0)}\right)=-n(m_1-m_0)+\ln\left(\frac{m_1}{m_0}\right)\sum_{i=1}^{n}x_i>\ln(k_\alpha)$$

$$\Leftrightarrow \quad \sum_{i=1}^{n}x_i>[\ln(k_\alpha)+n(m_1-m_0)]/\ln(m_1/m_0) \quad \text{car} \quad \ln(m_1/m_0)>0$$

$$\Leftrightarrow \quad \bar{x}>C$$

Ainsi la meilleure variable de décision est la moyenne $\overline{X}$ (ceci n'est pas une surprise étant donné que pour la loi de Poisson E(X)=m) et la région critique optimale est W={$\overline{X}$>C}. Ensuite, la construction du test se déroule de façon classique. L'échantillon est de taille n=40 (relevés quotidiens sur 2 mois), il est donc suffisamment grand pour approcher (grâce au TCL) la loi de $\overline{X}$ par une loi normale $N(\mu,\sigma^2/n)$ avec μ=E(X)=m et σ^2=V(X)=m. Ainsi sous l'hypothèse H_0, $\overline{X}$ suit une loi N(2,2/n), d'où

$$\alpha=P(W|H_0) \Leftrightarrow 0.05=P(\overline{X}>C) \Leftrightarrow 0.05=P\left(\sqrt{n}\frac{\overline{X}-2}{\sqrt{2}}>\sqrt{n}\frac{C-2}{\sqrt{2}}\right) \Leftrightarrow 0.05=P(Z>C')$$

où Z suit une loi N(0,1). A l'aide de la table de la loi normale, on obtient C'=1.64 puis C=2.37. Sous l'hypothèse H_1, $\overline{X}$ suit une loi N(5,5/n), d'où

$$1-\beta=P(W|H_1)=P(\overline{X}>C)=P\left(\sqrt{n}\frac{\overline{X}-5}{5}>\sqrt{n}\frac{C-5}{5}\right)=P(Z>-7.44)\approx 1\,.$$

- Si $\bar{x}>2.37$ alors on accepte H_1, *i.e* on considère la machine comme défaillante, avec 5% de chance de se tromper
- Si $\bar{x}<2.37$ alors on garde H_0, *i.e* on considère que la machine est en bon état, avec une chance de se tromper quasi nulle.

5. STATISTIQUE EXHAUSTIVE ET COMPLETE

5.1. Exhaustivité

Considérons une statistique $T=T(X_1,\ldots,X_n)$, comme par exemple un estimateur de θ. On dit que la statistique T est *exhaustive* pour θ si la probabilité conditionnelle conjointe de $(X_1,\ldots,X_n)$ sachant T ne dépend plus de θ.

Cette notion est très importante en statistique car elle signifie que toute l'information apportée par l'échantillon concernant θ est contenue dans T. Autrement dit, une fois T connue, aucune autre valeur de l'échantillon ou aucune autre statistique n'apportera de renseignements supplémentaires sur θ. Une statistique exhaustive exploite au maximum l'information contenue dans l'échantillon.

<u>Théorème de factorisation de Fisher-Neyman</u>
Notons $L(\mathbf{x};\theta)=L(x_1,\ldots,x_n;\theta)$ la fonction de vraisemblance de l'échantillon. Alors la statistique T est exhaustive pour θ si et seulement la fonction de vraisemblance peut s'écrire sous la forme

$$L(\mathbf{x};\theta)=g(T;\theta)\times h(\mathbf{x}),$$

où la fonction h ne dépend pas de θ et g est la loi de $T(X_1,\ldots,X_n)$.

<u>Exemple</u>
Supposons que l'échantillon suit une loi de Poisson $\mathcal{P}(\lambda)$, où λ est le paramètre à estimer. Alors la statistique

$$S=\sum_{i=1}^{n} X_i$$

est exhaustive. En effet, la fonction de vraisemblance est donnée par

$$L(x_1,\dots,x_n;\theta)=\prod_{i=1}^{n}\frac{e^{-\theta}\theta^{x_i}}{x_i!}=e^{-n\theta}(\theta^{\sum_{i=1}^{n}x_i})(\prod_{i=1}^{n}\frac{1}{x_i!})=\frac{s!}{n^s\prod_{i=1}^{n}x_i!}$$

$$=\underbrace{\frac{s!}{n^s\prod_{i=1}^{n}x_i!}}_{h(x)}\times\underbrace{\frac{e^{-n\theta}(n\theta)^s}{s!}}_{g(s;\theta)}$$

où h est indépendante de θ et g est la loi de S, *i.e* une loi de Poisson $\mathcal{P}(n\theta)$.
NB. On montre assez facilement que la somme de deux lois de Poisson indépendantes et une loi de Poisson en utilisant le binôme de Newton dans le développement de

$$P(X+Y=k)=P(\bigcup_{i=0}^{k}(X=i\cap Y=k-i))\,).$$

<u>Théorème de Rao-Blackwell</u>
Soit T un estimateur quelconque sans biais de θ et U une statistique exhaustive pour θ. Alors $T^*=E[T|U]$ est **un** meilleur estimateur sans biais de θ.

<u>Remarques</u>
- U est exhaustive donc la densité conditionnelle sachant U ne dépend pas de θ donc E[T|U] ne dépend pas de θ (toute l'information concernant θ est connue puisque U est connue).
- T^* est sans biais car d'après le théorème de l'espérance totale on a
$$E(T^*)=E(E[T|U])=E(T)=\theta$$
- T^* est meilleur que T car d'après le théorème de la variance totale on a
$$Var(T)=var(E[T|U])+E(var[T|U])=var(T^*)+E(var[T|U])\geq var(T^*).$$
- D'après le paragraphe 2.3., on sait que si de plus la variance de T^* est minimale (borne de Cramer-Rao) alors T^* est unique.

5.2. Complétude

On dit qu'une statistique $T=T(X_1,\dots,X_n)$ est *complète* pour une famille de lois de probabilités $f(x_1,\dots,x_n\,;\theta)$ si

$$\forall\theta,\ E[h(T)]=0 \Rightarrow h=0 \text{ (p.s.)}$$

<u>Exemple</u>
La statistique S est complète pour une loi de Poisson $P(\lambda)$. En effet,

$$E[h(S)] = \sum_{s=0}^{+\infty} h(s)P(S=s) = \sum_{s=0}^{+\infty} h(s)\frac{e^{-n\theta}(n\theta)^s}{s!} = e^{-n\theta}\sum_{s=0}^{+\infty} h(s)\frac{(n\theta)^s}{s!} = 0\,,\ \forall\theta$$

$$\Leftrightarrow\ \sum_{s=0}^{+\infty} h(s)\frac{(n\theta)^s}{s!} = 0\,,\ \forall\theta$$

Cette série ne peut être nulle pour tout θ que si h(s)=0 pour tout $s\in\mathbb{N}$.

Propriété de Lehman-Scheffé
Si T est un estimateur sans biais de θ dépendant d'une statistique exhaustive complète U, alors T est **l'unique** estimateur sans biais de variance minimale de θ.

En particulier, si T est un estimateur sans biais de θ alors $T^*=E[T|U]$ est le meilleur estimateur sans biais de θ.

Exemple
Soit X la variable aléatoire comptabilisant les commandes hebdomadaires d'un produit. On sait que X suit une loi de Poisson P(λ). On cherche à estimer la probabilité que X soit nul à partir d'observations sur *n* semaines.

- Paramètre à estimer : $P(X=0)=e^{-\theta}$
- Estimateur : Fréquence : $T=K/n$ où K est le nombre de semaines avec aucune commande.

$$E[T]=e^{-\theta} \text{ et } var(T)=e^{-2\theta}(e^{\theta}-1)/n$$

- Statistique exhaustive et complète : S
- Meilleur estimateur sans biais : $T^*=E[T|S]$

$$T^* = \left(1-\frac{1}{n}\right)^{n\bar{X}} \text{ et } var(T^*)=e^{-2\theta}(e^{\theta/n}-1)<var(T)$$