Concepts Fondamentaux de l'Estimation

Introduction

Nous nous intéressons à une population ou à un processus et obtenons des informations à son sujet en collectant des données sur un échantillon. La grande question est de savoir ce que notre échantillon particulier nous permet de déduire sur la population ou le processus qui nous intéresse. Pour répondre à cette question, nous avons d'abord besoin d'un modèle pour cette population ou ce processus.

Objectif de la statistique inférentielle ?

- étendre les propriétés trouvées sur un n-échantillon à la population totale.
- vérifier la validité de ces propriétés.
- Calculer l'erreur commise en étendant les propriétés à la population entière.

Quelques Exemples

EXEMPLE 1

Un ingénieur doit construire une digue destinée à protéger une ville contre d'éventuelles inondations résultant des crues d'une rivière. Du choix de la hauteur h découlent les conséquences économiques suivantes :

- Coût de construction de la digue est de $10h$.
- Les dommages occasionnés par les inondations pouvant survenir dès que la côte maximale X de la crue dépasse h :

$$\begin{cases} 0 \ \text{ si } \ X \leq h \\ 100(X-h) \ \text{ si } \ X > h. \end{cases}$$

Si on note f la densité de la v.a. X, le coût w pour l'ingénieur, résultant du choix de h est :

$$w = 10h + 100 \int_h^{+\infty} (x-h) f(x) dx.$$

$$\text{Si} \quad f(x) = \frac{1}{\theta} e^{-x/\theta} \mathbb{1}_{\mathbb{R}_+^*}(x) \quad \text{alors} \quad w(h,\theta) = 10h + 100\theta e^{-h/\theta}$$

Si θ est connue, la côte optimale est $h^* = \theta \ln 10$. Mais, généralement θ est inconnue.

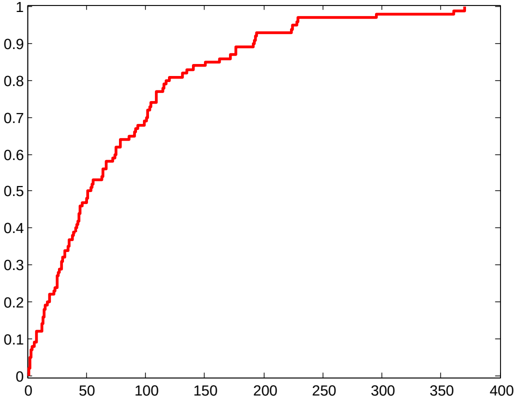
Exemple 2

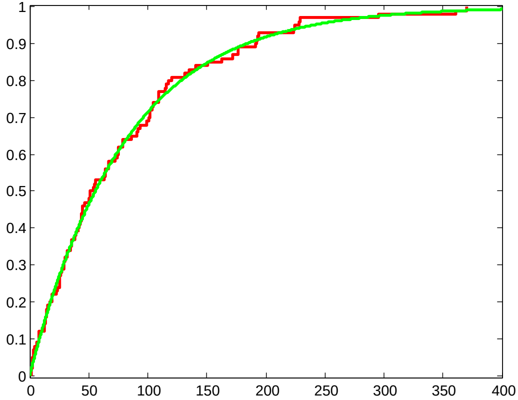
Dans le but d'étudier la densité du trafic sur internet, on a mesuré les durées de transfert, en millisecondes, d'un même message entre deux sites, à 100 moments différents d'une même journée :

31.76	67.23	29.17	108.76	370.36	50.44	102.01	4.42	7.85	74.77
295.57	31.86	14.21	2.31	229.27	63.82	43.23	30.07	26.3	34.45
25.51	78.71	43.36	172.25	228.26	1.63	100.47	41.53	3.08	134.65
131.58	85.91	5.55	75.19	103.79	150.73	73.88	78.52	162.53	66.58
191.30	38.93	360.94	44.51	18.88	115.72	64.23	45.05	223.22	56.04
35.41	120.67	14.22	2.34	13.34	54.49	102.15	23.37	115.19	3.48
38.20	,40.83	140.54	192.32	109.23	12.19	7.45	28.57	25.70	117.61
44.12	16.79	72.53	0.93	192.63	90.91	53.64	224.26	25.65	46.87
50.88	14.90	12.18	18.95	91.69	63.48	22.77	176.68	176.76	104.78
51.60	2.21	27.35	93.21	98.62	12.96	7.40	35.79	109.20	193.83

On souhaite connaître la durée moyenne de transfert, la probabilité qu'un transfert se fasse en moins de 10 ms ou en plus de 200 ms, etc...
Notons $x_1, \ldots, x_n$ $(n = 100)$ ces observations.
On peut alors se poser les questions suivantes :

- Au vu de ces observations, est-il raisonnable de supposer que la durée de transfert d'un message est une v.a. de loi exponentielle ?
- Comment proposer une valeur (ou un ensemble de valeurs) vraisemblable pour les paramètres de cette loi ?
- Que peut-on garantir aux usagers d'internet sur la durée de transfert des messages ? Sur un paquet de 100 messages, combien seront transférés en moins de 50 ms ?


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
50
100
150
200
250
300
350
400


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
50
100
150
200
250
300
350
400

Définition

Soit $\mathcal{P}$ une famille de lois de probabilité sur un espace probabilisable $(\mathcal{X},\ \mathcal{A})$. On appelle modèle statistique le triplet $(\mathcal{X},\ \mathcal{A},\ \mathcal{P})$.

$\mathcal{X}$ = L'espace des observations.
La famille $\mathcal{P}$ est souvent décrite à l'aide d'un paramètre.

Définition

On appelle modèle statistique paramétrique un modèle statistique $(\mathcal{X},\ \mathcal{A},\ \mathcal{P})$ tel que :

$$\exists p \in \mathbb{N}\ ;\ \ \mathcal{P} = \{P_\theta\ ;\ \theta \in \Theta \subset \mathbb{R}^p\}$$

Θ est appelé l'espace des paramètres.

Souvent on utilise le modèle d'échantillonnage où les X_i sont i.i.d. : on réserve alors la notation $\mathbb{L}_\theta$ à leur loi commune.
On note $X_1, X_2, \ldots X_n$ les v.a. donnant les résultats correspondants (de même loi de probabilité $\mathbb{L}_\theta$, indépendantes).

$(X_1, \ldots, X_n)$ forme un échantillon de taille n. Toute fonction mesurable de ces v.a. est appelée statistique.

Que peut-on dire sur la loi de probabilité $\mathbb{L}$?

- Comment évaluer la moyenne ou la variance de cette loi ? (estimation ponctuelle)
- Peut-on obtenir des intervalles de confiance de ces paramètres ? (estimation ensembliste)
- $\mathbb{L}$ appartient-elle à une famille de lois théoriques connues ? (tests d'ajustement).

Deux grandes parties se distinguent en statistique inférentielle :

$\mathbb{L} \in$ à une famille de lois paramétrées par $\theta \in \Theta$

statistique paramétrique

Si une telle hypothèse ne peut être faite

statistique non paramétrique

Un premier problème posé est celui de la recherche de la forme de la loi de probabilité $\mathbb{L}$.

Les outils les plus utilisés sont :

- Le diagramme des fréquences observées (cas discret) et l'histogramme (cas continu).
- La fonction de répartition empirique.

a) Histogramme

L'ensemble des observations $x_1, \ldots, x_n$ est réparti en k classes de type $[a_i, a_{i+1}[$, $i = 0, \ldots, k-1$. On trace les rectangles s'appuyant sur ces intervalles et de hauteur $h_i = \dfrac{1}{a_{i+1} - a_i} \dfrac{n_i}{n}$.

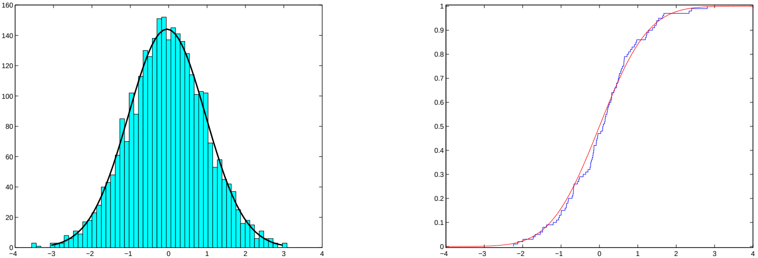
b) Fonction de répartition empirique

C'est un outil très utilisé lorsqu'on fait du "choix de modèle".

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

Th. *Glivenko-Cantelli*

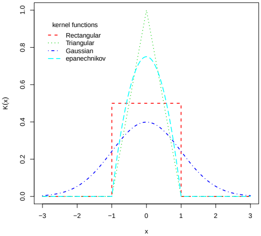
$$F_n \underset{n \to +\infty}{\longrightarrow} F$$

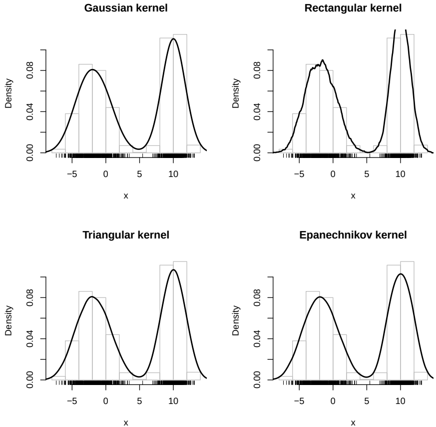


c) Méthode du noyau

$$\hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)$$

- Noyau uniforme $K(x) = \frac{1}{2} \mathbb{1}_{\{-1 \leq x \leq 1\}}$.
- Noyau gaussien $K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$.
- Noyau triangulaire $K(x) = \Big(1 - |x|\Big) \mathbb{1}_{\{-1 \leq x \leq 1\}}$.
- Noyau d'Epanechnikov $K(x) = \frac{3}{4} \max(1 - x^2, 0)$.






```
par(mfrow=c(2,2))
hist(x,  probability = TRUE, main = "Gaussian kernel", border = "gray")
lines(density(x,  window = "gaussian"), lwd = 2)

hist(x, probability = TRUE, main = "Rectangular kernel", border = "gray")
lines(density(x, width = 3, window = "rectangular"), lwd = 2)

hist(x,  probability = TRUE, main = "Triangular kernel", border = "gray")
lines(density(x,  window = "triangular"), lwd = 2)

hist(x,  probability = TRUE, main = "Epanechnikov kernel", border = "gray")
lines(density(x,  window = "epanechnikov"), lwd = 2)
```

Soit $(x_1, \ldots, x_n)$ une observation d'un échantillon $(X_1, \ldots, X_n)$, X_i de loi $\mathbb{L}(\theta)$, seul un paramètre θ restant inconnu.

PROBLÈME

Comment peut-on *"estimer"* θ à partir de n observations $x_1, x_2, \ldots, x_n$?

DÉFINITION

Un estimateur T (ou estimateur ponctuel) est une fonction de $(X_1, \ldots, X_n)$ que nous utilisons pour inférer la valeur d'un paramètre.

Cette question recouvre deux problèmes :

1. Définir un estimateur possédant de "bonnes qualités".
2. Trouver la manière adéquate de le choisir.

Définition

Une statistique T est appelée **estimateur sans biais** *de θ si $E_\theta(T) = \theta \quad (\forall\theta)$. Un estimateur qui n'est pas sans biais est dit biaisé et la différence $E_\theta(T) - \theta = b(T, \theta)$ est appelée le biais de l'estimateur.*

Définition

Un estimateur T de θ est dit **convergent** *si*

$$\forall\varepsilon > 0 \quad P(|T - \theta| \geq \varepsilon) \longrightarrow 0$$

Quand la taille de l'échantillon devient grande, la probabilité que l'estimation s'écarte de θ devient petite.

Un estimateur T sans biais dont la variance $V(T) \underset{n \to +\infty}{\longrightarrow} 0$ est convergent.

Remarque : L'espérance étant un paramètre de centrage la condition *sans biais* nous assure donc que T prendra ses valeurs autour de θ. La variance faible assure ensuite que la probabilité que T prenne une valeur très différente de θ est très faible.

Biais et transformation

- Les transformations non linéaires des estimateurs sans biais ne sont généralement pas sans biais : si T est un estimateur sans biais de θ, g(T) n'est généralement pas un estimateur sans biais de $g(\theta)$.
- Par exemple, si T est un estimateur sans biais de θ, alors T^2 n'est pas un estimateur sans biais de θ^2. En effet

$$E_\theta(T^2) = V(T) + (E_\theta(T))^2 = V(T) + \theta^2$$

- Donc, à moins que T soit concentré en un point (!), $V(T) > 0$ et T^2 est positivement biaisé, pour tout $\theta \in \Theta$, $b(T^2, \theta^2) = E_\theta[T^2] - \theta^2 > 0$. A l'inverse, si T^2 est un estimateur sans biais de θ^2, alors $|E_\theta[T]| < |\theta|$.

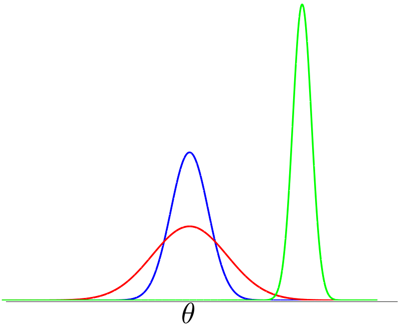
Erreur quadratique moyenne

- Pour mesurer la performance d'un estimateur, il est pertinent de calculer l'erreur quadratique moyenne (ERQM). l'ERQM mesure la dispersion de l'estimateur autour de la "vraie" valeur du paramètre.

$$ERQM(T) = E\left[(T-\theta)^2\right] = (b(T,\theta))^2 + V(T)$$

où $b(T,\theta)$ est le biais de l'estimateur et $V(T)$ est la variance.

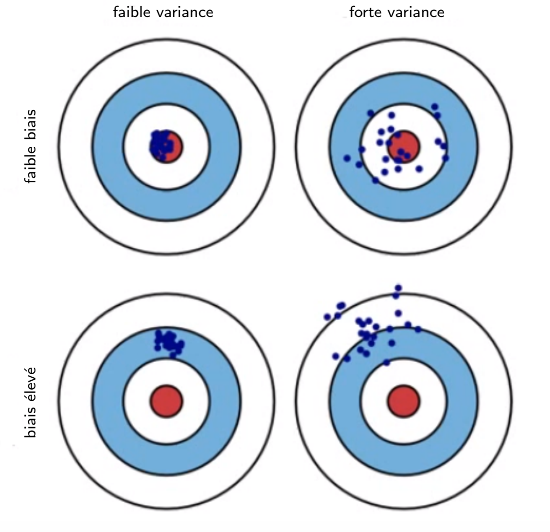
Soit T_1 et T_2 deux estimateurs de θ.
On dira que T_1 est plus efficace que T_2 si $\text{ERQM}(T_1) < \text{ERQM}(T_2)$
Si T est sans biais $\text{ERQM}(T) = \text{V}(T)$



Densités d'estimateurs du paramètre θ

Remarque :

* On peut aussi utiliser $E(|T-\theta|)$ ou $E(|T-\theta|^{\alpha})$, ce qui conduit à d'autres estimateurs que la moyenne empirique.


faible variance
forte variance
faible biais
biais élevé

Exemple

Soit $(X_1, \dots, X_n)$ un échantillon, X_i est de loi $\mathcal{U}_{[0,\theta]}$, θ étant inconnu. On considère les deux estimateurs suivants : $T_1 = \dfrac{2}{n}\displaystyle\sum_{i=1}^{n} X_i$ et $T_2 = \frac{n+1}{n} X_{(n)}$.

Montrer que T_2 est un meilleur estimateur de θ que T_1.

Définition

Un estimateur T de θ sera dit asymptotiquement sans biais si $E_\theta(T) \longrightarrow \theta$

1) Utilisation des résultats d'échantillonnage

Les propriétés d'un échantillon permettent de proposer des estimateurs simples et possédant de nombreuses propriétés intéressantes (sans biais, convergence, normalité asymptotique, ...)

- **Le paramètre μ est l'espérance**

  $\hat{\mu} = \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ est un estimateur sans biais et convergent de μ.

- **Le paramètre σ^2 est la variance**

  $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ est un estimateur sans biais et convergent de σ^2.

2) Méthode des moments

Cette méthode, qui généralise l'approche des résultats d'échantillonnage, repose sur le fait que les moments empiriques convergent vers les moments théoriques. S'il y a plusieurs paramètres à estimer, on les détermine en prenant autant d'équations de la forme :

moment de la v.a. = moment empirique correspondant de l'échantillon.

Définition

Soit un échantillon $(X_1, X_2, \ldots, X_n)$ de loi $L(\theta)$ où $\theta \in \Theta \subseteq \mathbb{R}^k$, et telle que pour tout $\theta \in \Theta$ il existe un moment $\mu_k(\theta)$ d'ordre k. Si, pour toute réalisation $(x_1, x_2, \ldots, x_n)$ de $(X_1, X_2, \ldots, X_n)$ le système à k équations

$$\begin{cases} \mu_1(\theta) & = m_1 \\ \quad \vdots & \\ \mu_k(\theta) & = m_k \end{cases}$$

(où $m_r = \frac{1}{n}\sum_{i=1}^{n} x_i^r$) admet une solution unique, cette solution est appelée estimation des moments de θ. La fonction (de $\mathbb{R}^n$ dans $\mathbb{R}^k$) qui à toute réalisation $(x_1, x_2, \ldots, x_n)$ fait correspondre cette solution définit, en s'appliquant à $(X_1, X_2, \ldots, X_n)$, une statistique à valeurs dans $\mathbb{R}^k$ appelée estimateur des moments de θ.

Exemples :

Loi gamma

Soit X une v.a. de loi gamma de paramètres r et a, son espérance et sa variance valent :

$$E(X) = \frac{r}{a} \quad \text{et} \quad V(X) = \frac{r}{a^2}.$$

On a alors :

$$r = \frac{\big(E(X)\big)^2}{V(X)} \quad \text{et} \quad a = \frac{E(X)}{V(X)}.$$

Si on dispose d'un échantillon $(X_1, X_2, \ldots, X_n)$ de le loi gamma de paramètres r et a, la moyenne empirique $\bar{X}_n$ et la variance empirique S_n^{*2} sont des estimateurs convergents de $E(X)$ et $V(X)$ respectivement. On en déduit des estimateurs convergents de r et a :

$$\hat{r} = \frac{\bar{X}_n^2}{S_n^{*2}} \quad \text{et} \quad \hat{a} = \frac{\bar{X}_n}{S_n^{*2}}.$$

Loi de Gumbel

Soit X une v.a. de loi de Gumbel de paramètres α et β, son espérance et sa variance valent :

$$E(X) = \alpha + \gamma\beta \quad \text{et} \quad V(X) = \frac{\pi^2\beta^2}{6} \quad \text{(où } \gamma \text{ la constante d'Euler).}$$

On résout :

$$\left\{ \begin{array}{ll} \alpha + \gamma\beta & = \bar{X}_n \\ (\alpha + \gamma\beta)^2 + \frac{\pi^2\beta^2}{6} & = \displaystyle\frac{1}{n}\sum_{i=1}^{n} X_i^2 \end{array} \right.$$

ou, de façon équivalente :

$$\left\{ \begin{array}{ll} \alpha + \gamma\beta & = \bar{X}_n \\ \frac{\pi^2\beta^2}{6} & = S_n^{*2} \end{array} \right.$$

Ce qui donne la solution :

$$\hat{\beta} = \frac{\sqrt{6}}{\pi} S_n^* \quad \text{et} \quad \hat{\alpha} = \bar{X}_n - \frac{\gamma\sqrt{6}}{\pi} S_n^*.$$

3) Méthode du Maximum de Vraisemblance

On appelle vraisemblance la fonction :

$$x = (x_1, \ldots, x_n) \longrightarrow L(x, \theta)$$

définie par :

$$L(x,\theta) = \begin{cases} P_\theta[(X_1, \ldots, X_n) = (x_1, \ldots, x_n)] = \prod_{i=1}^{n} P_\theta(X_i = x_i) \quad \text{(discret)} \\ f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_\theta(x_i) \quad \text{(cas continu)} \end{cases}$$

Définition

Pour une observation $x = (x_1, \dots, x_n)$ donnée, l'estimation du M.V. de θ est le réel $t(x_1, \dots, x_n)$, s'il existe, qui rend maximum la vraisemblance $L(x, \theta)$:

$$L(x, t(x)) = \sup_{\theta \in \Theta} L(x, \theta)$$

On appelle alors estimateur du M.V. la statistique $T(X_1, \dots, X_n)$ dont une valeur observée est $t(x_1, \dots, x_n)$.

Remarque : La probabilité et la densité utilisées dans cette définition sont des fonctions des observations $x_1, \dots, x_n$, dépendant du paramètre θ. A l'inverse, la fonction de vraisemblance est considérée comme une fonction de θ dépendant des observations $x_1, \dots, x_n$.

Exemple :

1) Soit $(X_1, \dots, X_n)$ un échantillon de v.a. de loi exponentielle de paramètre θ, θ étant inconnu. Déterminer l'estimateur du M.V. de θ.

2) Soit $(X_1, \dots, X_n)$ un échantillon de v.a. de loi $\mathcal{U}_{[0,\theta]}$, θ étant inconnu. Déterminer l'estimateur du M.V. de θ.

THÉORÈME *(d'invariance fonctionnelle)*

Soit g une fonction mesurable bijective de Θ sur un ouvert Λ de $\mathbb{R}^p$. T est un estimateur de M.V. de θ ssi $g(T)$ est un estimateur de M.V. de $\lambda = g(\theta)$.

Posons $h(x, \lambda) = L(x, g^{-1}(\lambda)), \quad \lambda \in \Lambda$ ou $h(x, g(\theta)) = L(x, \theta), \quad \theta \in \Theta$.

La condition

$$L(x, t(x)) = \sup_{\theta \in \Theta} L(x, \theta)$$

est équivalente à

$$h(x, g(t(x))) = \sup_{\theta \in \Theta} h(x, g(\theta))$$

ou encore à

$$h(x, g(t(x))) = \sup_{\lambda \in \Lambda} h(x, \lambda)$$

Exemple :

Soit $(X_1, \ldots, X_n)$ un échantillon de v.a. de loi $N(\mu, \sigma^2)$. Déterminer l'estimateur de M.V. de $\theta = (\mu, \sigma^2)$.

On suppose maintenant que X_i suit une loi Lognormale. Déterminer l'estimateur de M.V. de la moyenne a et de la variance d^2 de X_i.

• L'estimateur du maximum de vraisemblance de (μ, σ^2) est

$$\hat{\mu}_n = \bar{X}_n \quad \text{et} \quad \hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2.$$

• L'estimateur du maximum de vraisemblance de (a, d^2) est

$$\hat{a}_n = exp(\hat{\mu'}_n + \frac{\hat{\sigma'}_n^2}{2}) \quad \text{et} \quad \hat{d}_n^2 = \exp(2\hat{\mu'}_n + \hat{\sigma'}_n^2) \times (\exp \hat{\sigma'}_n^2 - 1)$$

où $\hat{\mu'}_n = \bar{Y}_n$ et $\hat{\sigma'}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2$ avec $Y_i = \ln X_i$.

Autre méthode d'estimation

Parmi les nombreuses méthodes introduites dans la littérature, citons : **la méthode des moindres carrés**

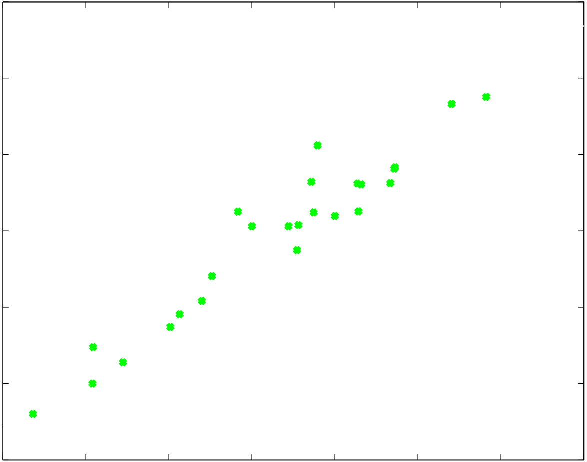
soit le modèle:

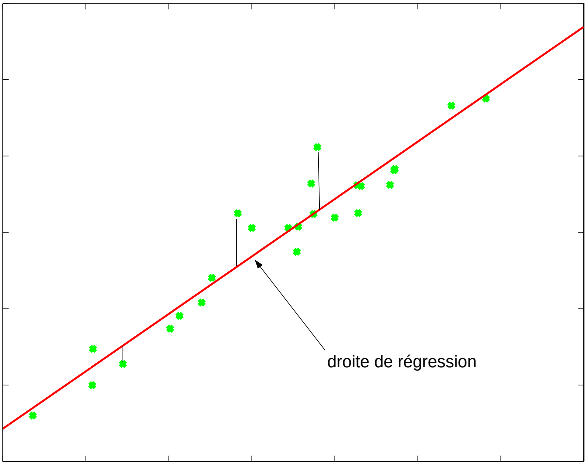
$$x_i = \phi(\underbrace{\theta_1, \ldots, \theta_k}_{\text{paramètres}} ; \underbrace{u_{i,1}, \ldots, u_{i,k}}_{\text{v. explicatives}}) + \underbrace{e_i}_{\text{résidu}} \qquad (i = 1, \ldots, n)$$

On estime alors les paramètres inconnus $\theta_1, \ldots, \theta_k$ en minimisant :

$$\min_{\theta_1, \ldots, \theta_k} \sum_{i=1}^{n} \left(x_i - \phi(\theta_1, \ldots, \theta_k ; u_{i,1}, \ldots, u_{i,k})\right)^2 .$$

Remarque : *Cette méthode est équivalente à la méthode du M.V. lorsque l'on suppose que les résidus sont normalement distribués.*




droite de régression

$$y_i = ax_i + b + e_i \quad \text{pour } i = 1, \ldots, n$$

Notons $EQ(a, b)$ la fonction de $\mathbb{R}^2$ dans $\mathbb{R}^+$ définie par :

$$EQ(a, b) = \sum_{i=1}^{n} (y_i - ax_i - b)^2.$$

La fonction admet un minimum pour

$$\hat{a}_n = \frac{c_{xy}}{s_x^2} \quad \text{et} \quad \hat{b}_n = \bar{y}_n - \hat{a}_n \bar{x}_n$$

Eléments de la théorie de l'information

Soit le modèle $(\mathcal{X}, \mathcal{A}, \{P_\theta,\ \theta \in \Theta\})$, $\Theta \subset \mathbb{R}$. On fait les hypothèses suivantes :

- H_1 : $\forall x$, $\forall \theta$ $f_\theta(x) > 0$
- H_2 : $\forall x$, $\forall \theta$ f_θ est dérivable au moins deux fois par rapport à θ
- H_3 : On peut dériver au moins deux fois $\displaystyle\int_A f_\theta(x)dx$ par rapport à θ sous le signe d'intégrale, $\forall A \in \mathcal{A}$.

Définition

On appelle quantité d'information de Fisher $I_n(\theta)$ apportée par un n-échantillon sur le paramètre θ la quantité positive ou nulle suivante :

$$I_n(\theta) = E[(\frac{\partial \ln L}{\partial \theta})^2]$$

Théorème

Sous H_1, H_2 et H_3 on a : $I_n(\theta) = -E(\frac{\partial^2 \ln L}{\partial \theta^2})$

Sous les hypothèses H_1, H_2 et H_3 on a :

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

Ceci veut dire que chaque observation a la même importance, ce qui n'est pas le cas pour la loi uniforme sur $[0, \theta]$ où il est évident que c'est la plus grande observation qui est la plus importante.

Soit $X \sim N(\mu, \sigma^2)$ où σ est connue. On a $I_1(\mu) = \frac{1}{\sigma^2}$, donc l'information apportée par une observation sur la moyenne est d'autant plus grande que la dispersion est petite.

Inégalité de Rao-Cramer

L'intérêt de la quantité d'information de Fisher est qu'elle fournit une borne inférieure pour la variance de n'importe quel estimateur sans biais de θ. Ce résultat s'exprime sous la forme de la propriété suivante :

Théorème

Sous H_1, H_2 et H_3 on a pour tout estimateur T sans biais de θ

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

Si T est un estimateur sans biais de $h(\theta)$

$$V(T) \geq \frac{(h'(\theta))^2}{I_n(\theta)}$$

Un estimateur qui atteint cette borne sera dit *estimateur efficace*

Exemple : Soit $(X_1, \ldots, X_n)$ un échantillon d'une loi $\mathcal{P}(\lambda)$.

$$
\begin{aligned}
\mathrm{V}(\bar{X}_n) &= \frac{\lambda}{n} \\
\mathrm{I}_1(\lambda) &= -E\left(\frac{\partial^2(-\lambda + X \ln \lambda)}{\partial \lambda^2}\right) \\
&= E\left(\frac{X}{\lambda^2}\right) = \frac{1}{\lambda}
\end{aligned}
$$

Donc $\bar{X}_n$ est l'estimateur sans biais et efficace de λ.

Exemple : Soit $(X_1, \ldots, X_n)$ un échantillon d'une loi $\mathcal{U}([0,\theta])$. Comme $f_\theta(x) = \frac{1}{\theta}$ pour $0 < x < \theta$,

$$E\left(\frac{\partial \ln L}{\partial \theta}\right)^2 = \frac{n^2}{\theta^2}$$

Et $X_{(n)} \sim f_n(x) = n\dfrac{x^{n-1}}{\theta^n}\mathbb{1}_{[0,\theta]}(x)$ implique

$$E\left(X_{(n)}\right) = \int_0^\theta \frac{n}{\theta^n}x^n dx = \frac{n}{n+1}\theta$$

donc

$$T(X_1, \ldots, X_n) = \frac{n+1}{n}X_{(n)}$$

est sans biais.
Mais

$$\mathrm{V}(T(X_1, \ldots, X_n)) = \frac{\theta^2}{n(n+2)} < \frac{\theta^2}{n^2} \qquad (!)$$

Borne inférieure = pas nécessairement atteinte

Exemple : Soit $(X_1, \ldots, X_n)$ un échantillon d'une loi $N(\mu, \sigma^2)$,

$$\frac{\partial^2}{\partial(\sigma^2)^2} \ln\left(\frac{1}{\sigma}\exp(-\frac{1}{2}(x-\mu)^2/\sigma^2)\right) = \frac{1}{2\sigma^4} - \frac{(x-\mu)^2}{\sigma^6}$$

et

$$E\left(\frac{\partial^2}{\partial(\sigma^2)^2} \ln f_{(\mu,\sigma)}(X)\right) = -\frac{1}{2\sigma^4}$$

donne la borne $\dfrac{2\sigma^4}{n}$ pour un estimateur sans biais de σ^2.

Mais

$$\mathrm{V}(S_n^2) = \frac{2\sigma^4}{n-1} > \frac{2\sigma^4}{n} \qquad (?)$$

Statistique exhaustive

On reprend le modèle de Bernoulli. Considérons un échantillon

$(X_1, \ldots, X_n)$ où $X_i \sim \mathcal{B}(\theta)$ et $T(X_1, \ldots, X_n) = \sum_{i=1}^{n} X_i$

Si on a $\underline{n = 5}$

$$x = (0, 0, 1, 1, 0) \longrightarrow T = 2$$

$$x = (1, 0, 0, 0, 1) \longrightarrow T = 2$$

En considérant T perd-t-on de l'information ?
On peut essayer de formaliser cette idée en examinant la loi conditionnelle de $(X_1, \ldots, X_n)$ sachant $T(X_1, \ldots, X_n) = t$.
Cette loi conditionnelle est indépendante de θ ; cette indépendance signifie intuitivement, que connaissant T, la connaissance de $(X_1, \ldots, X_n)$ n'apporte pas d'information supplémentaire sur θ.

Définition

Soit le modèle statistique $(\mathcal{X}, \mathcal{A}, \{P_\theta, \theta \in \Theta\})$. La statistique T sera dite exhaustive si la loi conditionnelle de X sachant $T(X) = t$ est indépendante du paramètre θ.

Théorème (**Théorème de factorisation de Fisher**)

Soit le modèle stat. $(\mathcal{X}, \mathcal{A}, \{P_\theta, \theta \in \Theta\})$. T *une stat.* $\mathcal{X}^n \longrightarrow \mathcal{Y}$. T *est exhaustive ssi* $\exists$ *une famille de fonctions mesurables* $\{g_\theta,\ \theta \in \Theta\}$ *de* $\mathcal{Y}$ *dans* $\mathbb{R}^+$ *et une fonction mesurable* h *de* $\mathcal{X}^n$ *dans* $\mathbb{R}^+$*, telles que* $L(x, \theta)$ *se met sous la forme :*

$$L(x, \theta) = g_\theta(T(x)).h(x)$$

Exemples :

1) Modèle de Bernoulli : $g_\theta(u) = \theta^u (1-\theta)^{n-u} \quad h \equiv 1$

$$T = \sum_{i=1}^{n} X_i \text{ est exhaustive pour } \theta.$$

2) Modèle de Poisson : $g_\theta(u) = \mathrm{e}^{-n\theta} \theta^u \quad h(x_1, \ldots, x_n) = \prod_{i=1}^{n} \frac{1}{x_i!}$

$$T = \sum_{i=1}^{n} X_i \text{ est exhaustive pour } \theta.$$

3) Modèle uniforme :

$$g_\theta(u) = \frac{1}{\theta^n} \mathbb{1}_{]-\infty,\theta]}(u) \quad h(x_1, \ldots, x_n) = \mathbb{1}_{[0,+\infty[}(\inf x_i)$$

$$T = \sup_{i=1,\ldots,n} X_i \text{ est exhaustive pour } \theta.$$

4) Modèle gaussien :

$$g_\theta(u, v) = \frac{1}{(\sigma\sqrt{2\pi})^n} \exp\left(-\frac{1}{2\sigma^2}(n(u-m)^2 + (n-1)v)\right)$$

$$T = \left(\frac{1}{n}\sum_{i=1}^{n} X_i, \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2\right) \text{ est exhaustive pour } \theta = (m, \sigma^2).$$

Application à l'estimation

On considère le modèle $(\mathcal{X}, \mathcal{A}, \{P_\theta, \theta \in \Theta\})$ et on s'intéresse à rechercher, à partir d'un estimateur sans biais de θ, un estimateur qui lui est préférable par réduction de la variance.

Théorème

Si $\exists$ un estimateur de θ sans biais, de variance minimale, il est unique p.s.

Théorème (Rao-Blackwell)

Soit T un estimateur sans biais de θ et U une statistique exhaustive. l'estimateur

$$h(U) = E(T|U)$$

est un estimateur sans biais de θ au moins aussi bon que T.

Corollaire

Si $\exists$ une statistique exhaustive U, alors l'estimateur T sans biais de θ de variance minimale ne dépend que de U

1. L'espérance
$$h(U) = E(T|U)$$
est bien un estimateur : espérance indépendante de θ grâce à l'exhaustivité

2. Méthode constructive d'amélioration :

(a) Partir d'un estimateur sans biais T_0
(b) définir $T_1(x) = E\left(T_0(X)|U(X) = U(x)\right)$

Définition

U est complète pour une famille de lois de probabilité P_θ, $\theta \in \Theta$, si

$$E_\theta(g(U)) = 0 \ \ \forall \theta \in \Theta \Longrightarrow g \overset{P_\theta^U}{=} 0 \ \ \forall \theta \in \Theta$$

Exemples

1) Modèle de Bernoulli (θ)

$$S = \sum_{i=1}^{n} X_i \text{ est complète.}$$

2) Modèle de Poisson (λ)

$$S = \sum_{i=1}^{n} X_i \text{ est complète.}$$

3) Modèle uniforme sur $[0, \theta]$

$$X_{(n)} = \sup_{i=1,\ldots,n} X_i \text{ est complète.}$$

4) Modèle exponentielle (θ)

$$S = \sum_{i=1}^{n} X_i \text{ est complète.}$$

Théorème (Lehmann-Scheffé)

Si T^* est un estimateur sans biais de θ dépendant d'une statistique exhaustive et complète U alors T^* est l'unique estimateur sans biais de variance minimale de θ.
Si T un estimateur sans biais de θ on a $T^* = h(U) = E(T/U)$.

Exemple

Le nombre de communications dans un intervalle de temps donné est une v. a. X qui suit une loi de Poisson $\mathcal{P}(\theta)$ où θ est inconnu. On cherche à estimer la probabilité que X soit nul. On note $X_1, \dots, X_n$ les observations de X pendant n jours. Le paramètre à estimer est $e^{-\theta} = P(X = 0)$. Donner l'estimateur sans biais de variance minimale de ce paramètre.