

STATISTIQUES

Les bases

Deuxième leçon

Ce matin

- Exemple d'estimation ponctuelle
- Introduction à l'estimation nonparamétrique
- Estimation d'un domaine de confiance

Exemple : lecture commentée d'un article de R. Barlow et Ch. Beeston

"Fitting using finite Monte Carlo samples" Computer Physics Com. 77 (1993) p.219-299

Position du problème : nos données nécessitent souvent l'estimation de la composition de l'échantillon considéré comme superposition de plusieurs sources, en général simulées par Monte Carlo.

ex: données = output d'un trigger "muon". Il y a une flopée de raisons pour qu'un événement ait un μ : ce μ peut provenir

- de la désintégration d'un Z, d'un W
- de la désintégration directe d'un b, ou en cascade, ou directe d'un c
- de la désintégration d'un K, d'un \bar{u} ...

On aimerait connaître les proportions de chaque source dans les données.

Bien évidemment, il n'existe pas de forme analytique des densités de proba. des sources en fonction des variables mesurées, donc les sources ont été produites par Monte-Carlo.

Il n'y a pas d'autre méthode que de transformer données et sources en histogramme (bien souvent multidimensionnel). Nous avons donc k boîtes, et les données deviennent un échantillon de k valeurs entières correspondant au nombre d'observations dans chaque boîte (d_i)

Ceci est fait également pour chacune des sources. (on suppose m sources)

Le modèle : soit n le nombre total d'événements (réels - données)

n_j , j de 1 à m , les tailles des échantillons simulés

désignons par f_i , i de 1 à k , les prédictions d'observations, qui dépendent bien entendu des m paramètres correspondant aux fractions de chaque source, et enfin par a_{ji} le nombre d'événements Monte Carlo de source j contenus dans le bin i .

clairement:
$$f_i = n \sum_{j=1}^m p_j a_{ji} / n_j \quad \left(n_j = \sum_{i=1}^k a_{ji} \right)$$

Les p_j sont des proportions, et doivent donc vérifier $\sum_1^m p_j = 1$, et sont les paramètres à estimer.

On est en plein problème d'ajustement d'un histogramme.

Il y a plusieurs approches possibles, toutes avec quelques inconvénients

1) Ce que j'ai appelé les "Moindres Carrés modifiés"
Ceci revient à minimiser la quantité $S^2 = \sum_i \frac{(d_i - f_i)^2}{d_i}$

En effet, les observations d_i sont des réalisations de variables de Poisson, dont on prend comme paramètre le d_i observé, en absence d'autre information plus fiable.
(Rappel: le dénominateur est bien la variance de la loi de Poisson associée au bin.)
estime

L'estimateur est correct (attention les auteurs le notent χ^2 alors que ce n'est pas un χ^2)
 On sait qu'il est convergent, vers un $\chi^2 (n-m)$ ddl, mais que la convergence est lente
~~Et c'est~~ De plus, l'estimation "brutale" du paramètre de Poisson peut se décaler,
 surtout (comme c'est souvent le cas) si des boîtes sont presque vides !

On peut améliorer cet estimateur en regroupant les boîtes de façon à ce qu'aucun d_i ne soit inférieur à 7.

2) Ce que j'ai appelé "Maximum de vraisemblance"

La loi de Poisson de chaque boîte s'écrit $e^{-f_i} \frac{f_i^{d_i}}{d_i!}$

La vraisemblance de l'histogramme est donc $L(\lambda) \propto \prod_i d_i! \lambda^{-\sum_i d_i} e^{-\sum_i f_i}$

Ceci permet d'éviter les regroupements ; on sait que la convergence est plus rapide ...

Problème commun aux deux approches :

La taille finie des échantillons MC pour les sources n'est pas prise en compte

Les fluctuations statistiques des a_{ji} ne sont pas prises en compte

Le facteur $\frac{n}{n_j}$ dans la formulation des f_i aide, mais il est rarement plus petit que $\frac{1}{10}$!

Comment tenir compte de cela ?

On peut imaginer trois méthodes

- 1) Dans les moindres carrés modifiés, mettre au dénominateur non pas la variance des d_i , mais la variance de la différence $d_i - f_i$ puisque f_i est lui aussi aléatoire par l'intermédiaire des a_{ji} . Il faut alors minimiser

$$S^2 = \sum_i \frac{(d_i - f_i)^2}{d_i + n^2 \sum_j a_{ji} / n_j^2}$$

OK, mais attention aux boîtes "presque" vides

- 2) Une solution non décrite par R. Barlow: utiliser non les a_{ji} , qui sont de mauvais estimateurs des densités de probabilité, mais passer à de meilleures estimations en "lissant" les histogrammes des MC. C'est la solution qui a été utilisée pour la recherche du Higgs à LEP! Pas facile!

- 3) R. Barlow modifie la formulation de la vraisemblance

V_j, V_i , le nombre (inconnu) d'événements attendus est A_{ji} , dont a_{ji} n'est qu'une réalisation. Par suite, la vraisemblance "totale" s'écrit

$$\ln \mathcal{L} \propto \sum_i d_i \ln f_i - f_i + \sum_i \sum_j a_{ji} \ln A_{ji} - A_{ji}$$

A priori très compliqué! Les auteurs ont trouvé des astuces pour simplifier et ont codé la méthode dans HBOOK (HMCMLL, ...)

Estimation non-paramétrique

- Il s'agit de trouver un estimateur de la ddp (inconnue) de la v.a. X sous jacente à l'échantillon $p(x)$
- Nous verrons successivement :
 - l'histogramme
 - l'estimateur de Rosenblatt
 - le $k^{\text{ième}}$ plus proche voisin

① L'histogramme:

c'est le plus courant des estimateurs non-paramétriques

il suffit de renormaliser et de prendre, dans le "bin" i

$$\hat{p}_i = \frac{n_i}{N}$$

← nombre d'observations dans le bin
← taille totale de l'échantillon.

si $g_i(x)$ désigne la fonction indicatrice du bin i

$$\begin{cases} g_i(x) = 1 & \text{si } x \in r_i \\ g_i(x) = 0 & \text{si } x \notin r_i \end{cases}$$

$$\hat{p}(x) = \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^N g_i(x_j) g_i(x) \quad (1)$$

Rappels:

le ddp des $\{n_i\}$ est la loi multinomiale, de paramètres p_i
(proba. d'avoir une observation dans le bin r_i lors d'un tirage isolé)

$$P[n_1, n_2, \dots, n_k] = N! \prod_{i=1}^k \frac{p_i^{n_i}}{n_i!} \quad ; \quad p_i = \int_{r_i} p(x) dx$$

← inconnu!

• Etude d'un "bin" :

La loi binomiale dit $E[n_i] = N p_i$; $V[n_i] = N p_i (1 - p_i)$

Comme nous avons $\hat{p}_i = n_i / N$, on en déduit

$$E[\hat{p}_i] = p_i \quad ; \quad V[\hat{p}_i] = \frac{1}{N} p_i (1 - p_i)$$

\hat{p}_i est un estimateur convergent non biaisé de p_i .

NB: si $p_i \ll 1$ (par ex si k grand) $V[n_i] \approx N p_i \approx N \hat{p}_i = n_i$

("règle" que vous utilisez tous les jours!)

• Corrélations de "bin" à "bin" :

$$\text{Cov}[n_i, n_j] = -N p_i p_j \quad (i \neq j) \quad \text{corrélation negative}$$

Si p_i et $p_j \ll 1$ corrélation est faible.

• L'histogramme globalement

Bien que \hat{p}_i soit un estimateur convergent de p_i ,

$\hat{p}(x)$ n'est pas un estimateur convergent de $p(x)$, à cause des discontinuités de $\hat{p}(x)$ aux bornes des "bins".

On mesure la variance d'une fonction aléatoire par

$$V[\hat{p}(x)] = E\left[\int (p(x) - \hat{p}(x))^2 dx\right]$$

par analogie avec $V[\hat{\theta}] = V[\hat{\theta} - \theta_0] = E[(\hat{\theta} - \theta_0)^2] - \underbrace{E[E[\hat{\theta} - \theta_0]]^2}_{b^2(\hat{\theta}) = 0}$

Les fonctions indicatrices g_i sont un cas particulier de fonctions orthogonales ψ_i

vérifiant donc $\int_{\mathbb{R}} \psi_i(x) \psi_j(x) dx = \delta_{ij}$

et $\hat{p}(x)$ est un développement sur ces fonctions $\hat{p}(x) = \sum_{i=1}^k c_i g_i(x)$

Si $p(x)$ était connu, le choix $c_i^* = \int_{\mathbb{R}} g_i(x) p(x) dx = E[g_i(x)]$

minimise la variance de $\hat{p}(x)$. Mais puisque $p(x)$ est inconnu, on estime

$$c_i^* \text{ par un } \hat{c}_i = \frac{1}{N} \sum_{j=1}^N g_i(x_j) \quad (\text{cf formule (1)})$$

$$V[\hat{p}(x)] = E\left[\int (p - \hat{p})^2 dx\right] \quad \text{j'vous laisse faire le calcul en exercice}$$

N.B: comme souvent, on remplace $p - \hat{p}$ par $(p - p^*) + (p^* - \hat{p})$

$$\text{où } p^*(x) = \sum_{i=1}^k c_i^* g_i(x)$$

On trouve rapidement

$$V[\hat{p}(x)] = \underbrace{E\left[\int (p - p^*)^2 dx\right]}_A + \underbrace{E\left[\int (p^* - \hat{p})^2 dx\right]}_B \quad \lim_{N \rightarrow \infty} B = 0$$

A est une partie indépendante de N, donc une constante systématique qui montre que $\hat{p}(x)$ n'est pas un estimateur convergent de $p(x)$.

2) Estimateur de Rosenblatt (ddp "naïve")

Soit $\underline{P}(x)$ la distribution cumulative empirique:

Les N observations sont classées par ordre croissant, et

$$\underline{P}(x) = \begin{cases} 0 & \text{si } x < x_1 \\ i/N & \text{si } x_i \leq x < x_{i+1} \\ 1 & \text{si } x \geq x_N \end{cases}$$

Une formulation du TCL

$$\lim_{N \rightarrow \infty} \underline{P}(x) = F(x) \quad (\text{fonction de répartition})$$

Rosenblatt a choisi (étudié) $\hat{p}(x) = \frac{1}{2h} \left(\mathbb{F}(x+h) - \mathbb{F}(x-h) \right)$

(nombre d'observations dans une fenêtre glissante de longueur $2h$, divisé par $2h$)

- $E[\hat{p}] = \frac{1}{2h} \left(\mathbb{F}(x+h) - \mathbb{F}(x-h) \right) = p(x) + \frac{h^2}{6} p''(x) + o(h^4)$

$\hat{p}(x)$ estime donc p avec un biais en h^2

- $V\left[\frac{\hat{p}}{\sqrt{N}}\right] = \frac{p(x)}{2hN} + \frac{h^4}{36} |p''(x)|^2 + o\left(\frac{1}{hN} + h^4\right)$

si l'on définit l'asymptotique ici comme $h \rightarrow 0$ et $hN \rightarrow \infty$

alors $V[\hat{p}] \rightarrow 0$

\hat{p} est convergent

- Le choix de h est libre ; on peut minimiser soit le biais, soit la variance

Si l'on paramétrise $h = cN^\alpha$ (pour assurer $h \rightarrow 0$ et $hN \rightarrow \infty$)

$\alpha = 1/5$ minimise la variance, mais c dépend de $p(x)$ et de $p''(x)$

- N.B. L'estimateur de Rosenblatt est un cas particulier des estimateurs de Parzen.

③ $k =$ ^{ième} plus proche voisin

L'estimateur de Rosenblatt présentait des progrès certains.

Néanmoins, idéalement, il aurait fallu que $h(N)$ varie en fonction de x !

Qu'a cela ne tiennne !

Soit $k(N)$ un entier fixé, et soit $h(N)$ la "distance" de x à son $k =$ ^{ième} plus proche voisin ; $h(N)$ devient dépendant de x . Et comme le nombre d'observations autour de x dans l'intervalle $[x-h; x+h]$ est k par définition

$$\hat{p}(x) = \frac{k}{2Nh} \quad \text{est un estimateur de } p(x)$$

Le choix de k est un peu similaire au choix du h de Rosenblatt

On note que
$$E[\hat{p}] = p(x) + \frac{1}{24} \frac{k^2}{N^2} \frac{p''(x)}{p^2(x)} + \dots$$

et
$$V[\hat{p}] = \frac{p^2(x)}{k} + \left(\frac{1}{24} \frac{p''(x)}{p^2(x)} \frac{k^2}{N^2} \right)^2 + \dots$$

La convergence est assurée si $\lim_{N \rightarrow \infty} k = \infty$ et $\lim_{N \rightarrow \infty} \frac{k}{N} = 0$

Si $p''(x)$ est petit ($p(x)$ sans variations brutales) ; k grand augmente la précision

Si $p''(x)$ est grand (près des pics) k grand augmente le biais !

De $V[\hat{p}(x)] \simeq \frac{p^2(x)}{k}$, on déduit que $\sigma[\hat{p}_N(x)] \simeq \hat{p}_N(x) / \sqrt{k_N}$

Donc l'erreur statistique de cet estimateur est proportionnelle à la densité (et non plus à la racine carrée de la densité, comme dans le cas des estimateurs à intervalle fixe)

De plus $\sigma(\hat{p}(x)) / \hat{p}(x) = \frac{1}{\sqrt{k}}$ est constant, indép. de x

En résumé :

Histogramme : le moins précis

Rosenblatt (ou Parzen) et voisins proches : plus précis

Voisins : présente l'avantage d'adaptabilité aux données.

Exemple numérique : sur un échantillon de taille $N = 710$ tiré par Monte-Carlo

selon la loi
$$p(x) = \frac{1}{2 \times 710} \frac{1}{\pi} \left[\frac{2}{1 + 2(x-0,5)^2} + \frac{20}{1 + 20(x-0,2)^2} \right]$$
 pour $x \in [0, 1]$

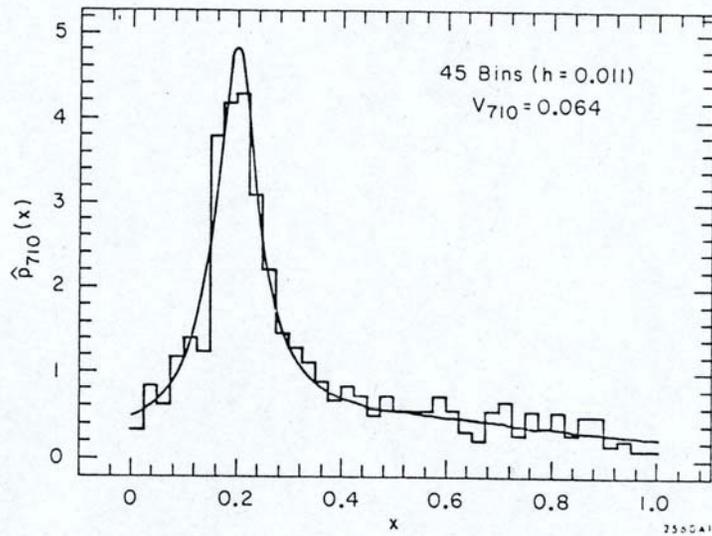


FIG. 1a Histogram density estimate.

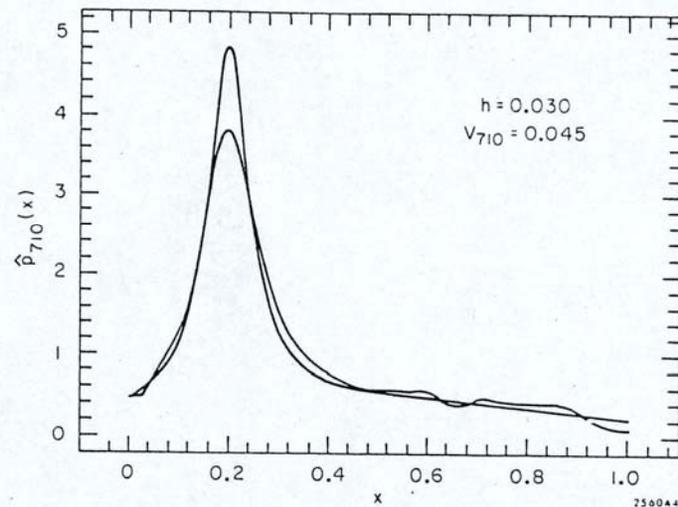


FIG. 1c Parzen (normal kernel) density estimate.

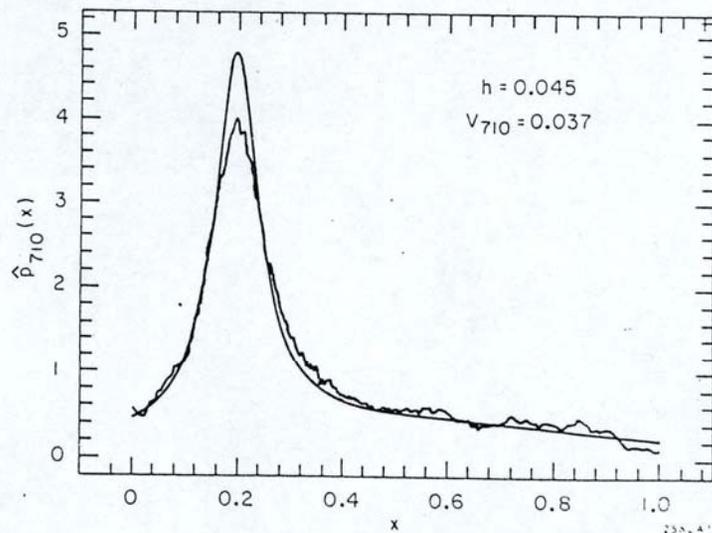


FIG. 1b Rosenblatt density estimate.

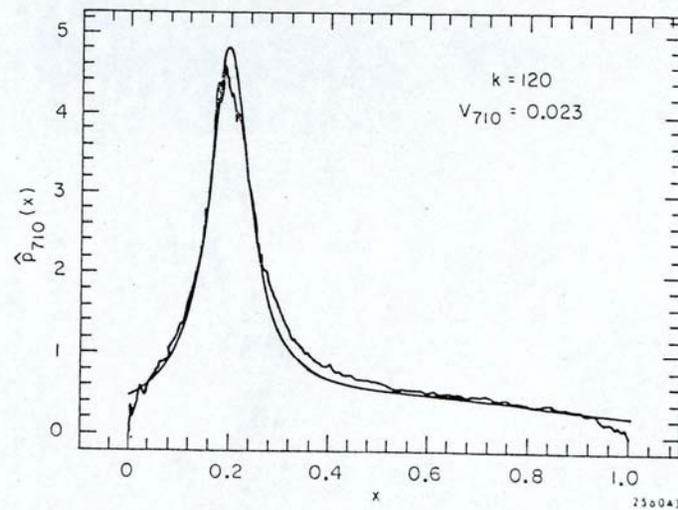


FIG. 1d k-th nearest neighbor density estimate.

Estimation d'un domaine de confiance

Estimation "par intervalle" (d'une région de confiance)

- L'estimation ponctuelle est clairement insuffisante
- θ scalaire : on veut trouver un intervalle $[\theta_a; \theta_b]$ tel que

$$\mathbb{P}(\theta_0 \in [\theta_a; \theta_b]) = \beta \quad (\text{parfois } \geq \beta)$$

intervalle de confiance

niveau de confiance

$\alpha = 1 - \beta$: risque d'erreur.

- Exemple : soit un n échantillon, v.a. X normale $\mathcal{N}(\theta_0, \sigma^2)$; σ^2 connu!

on sait que $\bar{x} = \frac{1}{n} \sum x_i$ suit $\mathcal{N}(\theta_0, \sigma^2/n)$

par suite : $z = (\bar{x} - \theta_0) / \sigma/\sqrt{n}$ suit $\mathcal{N}(0, 1)$

la table de $\mathcal{N}(0, 1)$ nous dit $\mathbb{P}(z \in [-1,96; 1,96]) = 0,95$

$$-1,96 \leq z \leq 1,96 \iff \underbrace{\bar{x} - \frac{1,96\sigma}{\sqrt{n}}}_{\theta_a} \leq \theta_0 \leq \underbrace{\bar{x} + \frac{1,96\sigma}{\sqrt{n}}}_{\theta_b}$$

• A noter

- 1) Les bornes θ_a et θ_b sont des statistiques, donc des v.a. (au moins une des bornes)
- 2) A β fixé, il existe une infinité d'intervalles de confiance!
 - si l'on "répartit le risque", on parle d'intervalle central.
 - on utilise aussi : la limite supérieure (ou inférieure)
ou l'"approche unifiée" de Feldman - Cousins.

- 3) Dans notre exemple, nous avons réussi parce que nous avons trouvé une statistique (z) dont la ddp. est indépendante du paramètre θ
(z est alors fonction de θ , pas sa ddp)!

Ceci nous a permis "d'inverser" et d'écrire

$$\beta = \mathbb{P}(a \leq z < b) = \mathbb{P}(\theta_a \leq \theta \leq \theta_b)$$

On a su trouver de telles statistiques dans quelques cas particuliers

- Loi parente normale : intervalle pour μ (σ^2 connu ou non)
intervalle pour σ^2 (μ quelconque)
- Loi parente Poisson : intervalle pour λ
- Loi parente Binomiale : intervalle pour p .

Méthode générale : la construction de Neyman

Soit t un estimateur de θ , de ddp connue $g(t|\theta)$
Fixons β Nous voulons que $\beta = \int_{t_1}^{t_2} g(t|\theta) dt$

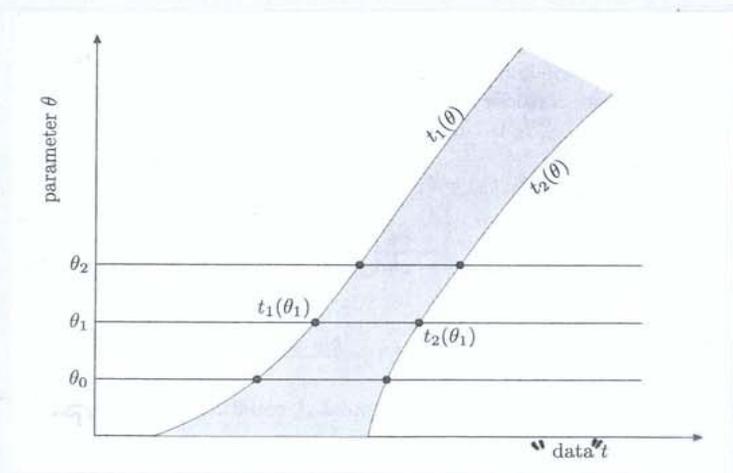
De façon à ce que la solution de cette équation intégrale soit unique,

ajoutons par ex. la condition pour un intervalle central

$$\frac{1-\beta}{2} = \int_{-\infty}^{t_1} g(t|\theta) dt = \int_{t_2}^{+\infty} g(t|\theta) dt$$

Pour chaque valeur de θ , nous pouvons ainsi calculer t_1 et t_2 , donc construire les courbes $t_1(\theta)$ et $t_2(\theta)$. La région (hachurée) entre ces deux courbes s'appelle l'enveloppe de confiance (confidence belt).

Si on observe une valeur t_{exp} à partir des données, l'intervalle de confiance pour θ sera donné par les intersections des courbes $t_1(\theta)$ et $t_2(\theta)$ avec la droite (v verticale) $t = t_{\text{exp}}$



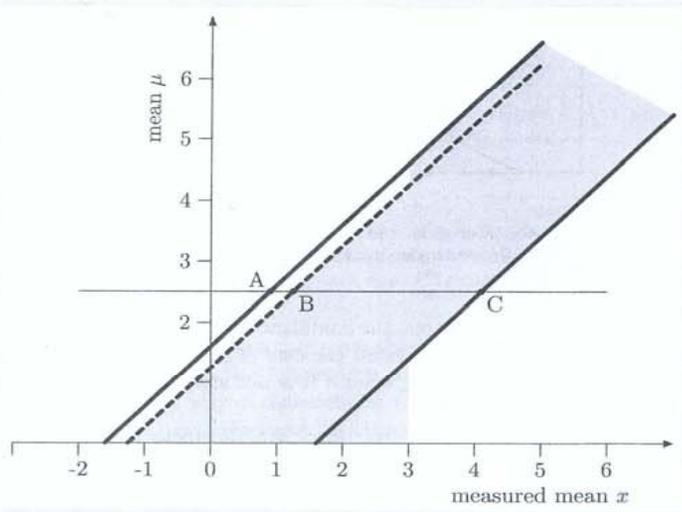
- La notion de couverture (coverage)

Toute méthode qui conduit à un intervalle $[\theta_a; \theta_b]$ tel que

$$\underline{P}(\theta_0 \in [\theta_a; \theta_b]) = \beta \quad \text{et dite posséder la propriété de couverture}$$

C'est le cas de la construction de Neyman.

- Limite supérieure : danger de flip-flopping



La limite supérieure est très utilisée pour donner un intervalle de confiance pour un paramètre qui ne peut pas être négatif mais probablement voisin de 0. (masse, ...)

Prenez l'exemple de la moyenne d'une loi Normale de variance σ^2 connue. La figure représente deux enveloppes de confiance ($\bar{\alpha} \beta = 90\%$). Entre les deux traits pleins, "intervalle central", alors que le trait pointillé donne la limite supérieure : la fonction $t_2(\theta)$ et alors à l'infini. Les deux enveloppes ont chacune la propriété de couverture.

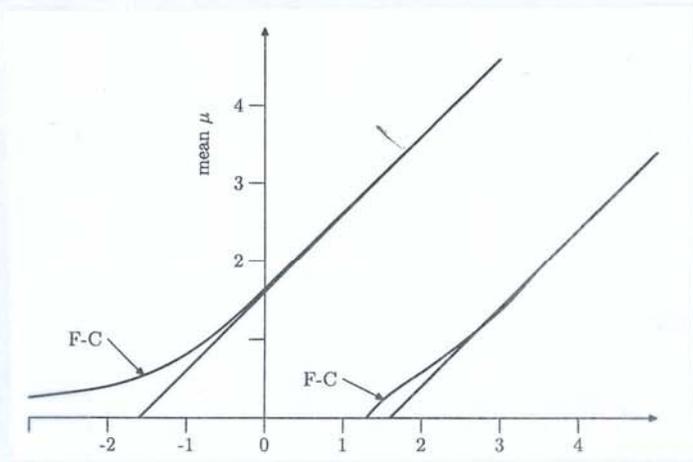
La zone hachurée représente l'enveloppe qui correspondrait à publier une limite sup si la mesure est à moins que 3 σ , et l'intervalle central

sinon. Une telle procédure est à proscrire : elle n'a pas la propriété de couverture (ex : si $\mu = 2.5$)

• L'approche unifiée de Feldman-Cousins

Pour déterminer l'intervalle pour $\mu = \mu_0$, F.C. accumule les "éléments" de probabilité $\underline{P}(x|\mu_0)$ dans l'ordre décroissant des rapport des vraisemblances

$$R(x) = \frac{\underline{P}(x|\mu_0)}{\underline{P}(x|\hat{\mu})} \quad \text{où } \hat{\mu} \text{ maximise } \underline{P} \text{ dans la région physique}$$



Exemple sur la moyenne : les deux enveloppes de confiance (F.C. et "central") coïncident pour $\mu > 1.65$

Élimination du danger de flip-flopping

Full coverage

Mais l'enveloppe dépend de l'échantillon $\{x\}$

- Données discrètes : l'équation intégrale de la construction de Neyman doit être remplacée par une somme qui n'atteint pas forcément β .

$$\int_{t_1}^{t_2} g(t|\theta) dt = \beta \quad \longrightarrow \quad \sum_{i=a}^b \underline{P}(t_i|\theta) \geq \beta$$

Il faut accepter un peu de "overcoverage"

Utilisation de la fonction de vraisemblance

- Soit (encore une fois) une ddp $\mathcal{N}(\mu, \sigma^2)$: μ inconnu, σ^2 connu
De la forme de la loi normale, on déduit que la log-vraisemblance est une parabole (en μ)

$$\ln \ell(\underline{x}, \mu) = C - \frac{(\mu - \underline{x})^2}{2\sigma^2}$$

On retrouve le maximum pour $\underline{x} = \mu$

Mais la ligne $\max \ell - \frac{1}{2}$ donne l'intervalle

$$\underline{x} - \sigma \leq \mu \leq \underline{x} + \sigma$$

soit l'intervalle à 68.3% de confiance.

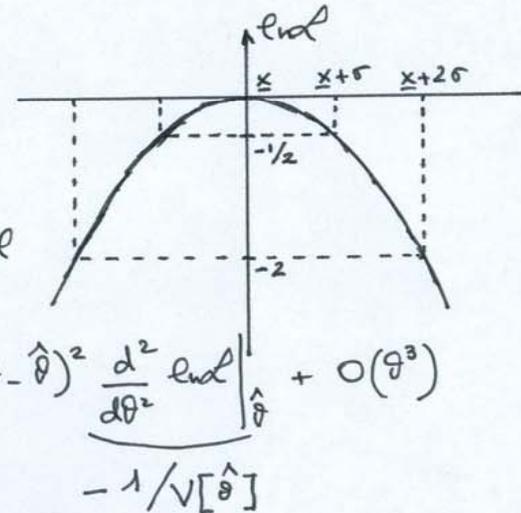
- Cette propriété est en fait utilisable \forall ddp, puisqu'un estimateur MV est asymptotiquement normal

Si nous faisons un développement limité autour de $\hat{\theta}$

$$\ln \ell(\theta) = \underbrace{\ln \ell(\hat{\theta})}_{0!} + (\theta - \hat{\theta}) \underbrace{\frac{d \ln \ell}{d \theta} \Big|_{\hat{\theta}}}_{0!} + \frac{1}{2} (\theta - \hat{\theta})^2 \underbrace{\frac{d^2 \ln \ell}{d \theta^2} \Big|_{\hat{\theta}}}_{-1/V[\hat{\theta}]} + o(\theta^3)$$

on retrouve la parabole!

- A échantillon fini, on n'a pas toujours une parabole! On peut
 - soit donner un intervalle de confiance dissymétrique par rapport à $\hat{\theta}$
 - soit trouver un changement de variable $\hat{\theta} \rightarrow \psi(\hat{\theta})$ qui "rapproche" de la parabole.



Utilisation d'approximations asymptotiques :

1°) Normalité asymptotique de l'estimateur MV
c'est en fait ce que l'on a utilisé ci-dessus

$$\hat{\theta} \xrightarrow[n \rightarrow \infty]{} \mathcal{N}(\theta_0, 1/n I_1(\theta_0))$$

où I_1 est la quantité d'information pour θ
contenue dans 1 observation.

on estime $I_1(\theta_0)$ par $I_1(\hat{\theta})$. La convergence est lente !

2°) Normalité asymptotique de $\frac{\partial \ln \mathcal{L}}{\partial \theta}$

$\forall n$, la statistique $T = \frac{\partial}{\partial \theta} \ln \mathcal{L}(x|\theta)$ suit $\mathcal{N}(0, n I_1(\theta))$

L'approximation ici consiste seulement à estimer $I_1(\theta)$ par $I_1(\hat{\theta})$
Méthode plus précise, convergence plus rapide, mais souvent difficile...

3°) Le rapport des vraisemblances

asymptotiquement $Q^2 = 2(\ln \mathcal{L}(x|\hat{\theta}) - \ln \mathcal{L}(x|\theta))$

suit une distribution de χ^2 à k ddl (k est la dimension de θ)

Là aussi, convergence lente, mais généralisable à un domaine de confiance.

Recherche d'une région (ou domaine) de confiance (θ multidimensionnel)

Nous avons déjà vu le cas de la vraisemblance et d'estimateurs MV

Quid des estimateurs par moindres carrés ?

Plaçons nous dans le cas du modèle linéaire et d'observations multivariées

L'estimation optimale de $\hat{\theta}$ se fait en minimisant (par rapport à θ)

$$S^2(\theta) = (\widetilde{y - A\theta}) V^{-1} (y - A\theta) \quad (1)$$

y les observations (échantillon)
 $A\theta$ modèle ("vraies valeurs" de y)
 V matrice de var. des y_i

Le minimum de S^2 est obtenu pour $\hat{\theta} = (\widetilde{A V^{-1} A})^{-1} \widetilde{A V^{-1} y}$
et $V[\hat{\theta}] = (\widetilde{A V^{-1} A})^{-1}$

Remplaçons dans (1) : il vient

$$S^2(\theta) = S^2(\hat{\theta}) + \underbrace{(\widetilde{\theta - \hat{\theta}}) V^{-1} [\hat{\theta}]}_{\chi^2} (\theta - \hat{\theta}) \quad (2)$$

$\Rightarrow y$ multivarié
 $\hat{\theta}$ multivarié

Chacun des 3 termes de (2) est un χ^2 , et nous avons n observations (y) indépendantes

et k paramètres

$S^2(\theta)$ est un χ^2 à n ddl

$S^2(\hat{\theta}) = S^2_{\min}$ est un χ^2 à $(n-k)$ ddl

$\Rightarrow X^2$ est un χ^2 à $[n - (n-k)] = k$ ddl

La statistique $X^2 = (\theta - \hat{\theta}) V^{-1}[\hat{\theta}] (\theta - \hat{\theta})$ ayant une distribution indépendante de θ a les bonnes propriétés pour obtenir une région de confiance autour de $\hat{\theta}$.

On coupe la surface $S^2(\theta)$ par des (hyper)plans d'équations $S_{\min}^2 + m$, et l'on obtient directement les régions de confiance

- cas unidimensionnel (θ scalaire) X^2 est un χ^2 à 1 dde

| | | | |
|---------------------|-----------------------|-------|--------------------------|
| $m = 1 \Rightarrow$ | région de confiance à | 68,3% | (parabus, à 1 σ) |
| $m = 4 \Rightarrow$ | " | 95,4% | — 2 σ |
| $m = 9 \Rightarrow$ | " | 99,7% | — 3 σ |
- cas bidimensionnel (θ à 2 composants) X^2 est un χ^2 à 2 dde

| | | |
|---------------------|---|-------|
| $m = 1 \Rightarrow$ | " | 39,3% |
| $m = 4 \Rightarrow$ | " | 86,5% |
| $m = 9 \Rightarrow$ | " | 98,9% |

• cas général

$\beta_m = \int_0^m \chi_k^2 d\chi_k^2$
 confiance associée à la valeur m de coupure \leftarrow
 \leftarrow ddp d'un χ^2 à k dde

Approche bayésienne:

- en approche fréquentiste, θ a une vraie valeur (θ_0) inconnue, ce n'est pas une v.a.
- en approche bayésienne

θ est une v.a. dont la distribution a posteriori $\pi(\theta|x)$ est obtenue par application du théorème de Bayes dès que l'on a

- un échantillon (représenté par x)
- une "croyance a priori" (prior belief) $\pi(\theta)$

$$\pi(\theta|x) = \mathcal{L}(x|\theta)\pi(\theta) / \int \mathcal{L}(x|\vartheta)\pi(\vartheta) d\vartheta$$

L'intervalle bayésien pour θ sera donc défini par θ_1 et θ_2 tels que

$$\int_{\theta_1}^{\theta_2} \pi(\theta|x) d\theta = \beta$$

Ici aussi, il y a une infinité de solutions : on le rend unique de différentes façons

1) $|\theta_2 - \theta_1|$ minimum

2) $\pi(\theta_1, x) = \pi(\theta_2, x)$

3) $\int_{-\infty}^{\theta_1} \pi(\theta|x) d\theta = (1-\beta)/2$
...

On ne peut pas parler de "couverture" pour de tels intervalles (il n'y a pas d'enveloppe de confiance). C'est pourquoi ces intervalles sont parfois appelés "credible intervals".