Estimation de paramètres

Christophe BLANC
LASMEA
Université Blaise Pascal
www.christophe-blanc.info

11 janvier 2009

Une représentation correcte de l'environnement routier à l'avant d'un véhicule repose, entre autre, sur la connaissance de l'état des différents usagers de la route. La détermination de ces différents états débute par l'estimation descriptive de ces états. On parlera de paramètres pour des quantités (scalaires / vecteurs) supposées temporellement invariantes. On cherchera par exemple à estimer la trajectoire des obstacles. Ces trajectoires seront définies par différents paramètres déterministes à évaluer. Les méthodes développées dans cette partie ne sont pour la plupart pas adaptées à une estimation temps réel des paramètres. Cependant, elles jugent de la capacité des modèles et des observations à estimer les paramètres déterminants pour une appréhension correcte de l'environnement routier.

Les méthodes statistiques d'estimation de paramètres sont basées sur la connaissance des composantes suivantes :

- l'ensemble où le(s) paramètre(s) à estimer, θ, prennent valeurs : l'espace de paramètres, Θ ;
- la loi de probabilité qui décrit l'effet du paramètre sur les observations : $p(Z/\theta)$ (paramètres déterministes : approche non bayésienne) ;
- la loi de probabilité qui décrit l'effet du paramètre sur les observations : $p(Z/\theta)$, et la densité de probabilité à priori de θ : $p(\theta)$ (paramètres aléatoires : approche bayésienne) ;
- l'ensemble où les observations, Z, prennent valeurs : l'espace des observations Z

Il existe donc deux modèles pour l'estimation de paramètres selon le caractère aléatoire de ceux ci. On parlera d'approche non bayésienne pour les paramètres déterministes. Dans ce cas, on utilise la densité de probabilité conditionnelle des mesures pour chaque valeur possible du paramètre : c'est la fonction de vraisemblance $p(Z/\theta)$. Quand les paramètres sont aléatoires, on parlera d'approche bayésienne. Pour ce deuxième modèle, on utilise la densité de probabilité à priori du paramètre à partir de laquelle on peut obtenir, par la formule de Bayes, la pdf a posteriori du paramètre :

$$p(\theta/Z) = \frac{p(Z/\theta)p(\theta)}{p(Z)} = \frac{1}{c}p(Z/\theta)p(\theta) \quad (1)$$

où c est une constante de normalisation qui ne dépend pas de θ.
Le problème d'estimation de paramètres est décrit ci-dessous. Etant donné les mesures :

$$z(j) = h[j, \theta, w(j)]_{j=1,\dots,k} \quad (2)$$

effectuées en présence de bruit $w(j)$, on cherche une fonction des k observations

$$\hat{\theta}(k) \triangleq \hat{\theta}[k, Z^k] \quad (3)$$

où les observations sont définies d'une manière compacte par :

$$Z^k = \{z(j)\}_{j=1,\dots,k} \quad (4)$$

qui estime la valeur de θ. La fonction 3 est appelée estimateur et sa valeur est l'estimé.

1 Les estimateurs

1.1 Estimateurs à maximum de vraisemblance et à maximum a posteriori

Une méthode commune d'estimation de paramètres non aléatoires est la méthode du maximum de vraisemblance[1] qui maximise la fonction de vraisemblance :

$$\hat{\theta}^{ML} = \arg\max_{\theta} p(Z/\theta) \quad (5)$$

L'estimateur MLE[2] est la solution de l'équation :

$$\frac{dp(Z/\theta)}{d\theta} = 0 \quad (6)$$

L'estimateur correspondant pour un paramètre aléatoire est l'estimateur à maximum a posteriori[3] qui est issu de la maximalisation de la pdf à posteriori :

[1] ML : Maximum Likelihood
[2] Maximum Likelihood Estimator
[3] MAP : Maximum A Posteriori

$$\hat{\theta}^{MAP} = \arg\max_{\theta} p(\theta/Z) = \arg\max_{\theta}[p(Z/\theta)p(\theta)] \quad (7)$$

1.2 Estimation des moindres carrés et minimisation du carré des erreurs

Une autre méthode commune d'estimation de paramètres déterministes est la méthode des moindres carrés[4]. Etant donné des mesures (scalaire, linéaire ou non-linéaire) :

$$z(j) = h(j, \theta) + w(j) \quad (8)$$

l'estimateur des moindres carrés de θ est :

$$\hat{\theta}^{LS}(k) = \arg\min_{\theta}\{\sum_{j=1}^{k}[z(j) - h(j,\theta)]^2\} \quad (9)$$

L'estimateur correspondant pour des paramètres aléatoires est l'estimateur qui minimise le carré des erreurs[5] :

$$\hat{\theta}^{MMSE}(Z) = \arg\min_{\hat{\theta}} E[(\hat{\theta} - \theta)^2/Z] \quad (10)$$

La solution de cette équation est l'espérance conditionnelle de θ :

$$\hat{\theta}^{MMSE}(Z) = E(\theta/Z) \triangleq \int_{-\infty}^{+\infty} \theta p(\theta/Z) d\theta \quad (11)$$

Tous ces estimateurs sont comparés dans [2]. Si, pour un ensemble de mesures données, les erreurs sont de moyennes nulles, gaussiennes et indépendantes alors l'estimé LS correspond à l'estimé ML. De même, l'estimé MAP d'une variable aléatoire gaussienne correspond à l'estimé MMSE. Pour évaluer la qualité des résultats, différentes méthodes sont détaillées dans la suite. On parlera d'estimateur non biaisé, de variance d'estimateur. De plus, la "limite d'information" sera appréhender par l'utilisation de la borne de Cramer-Rao et de l'information de Fisher.

1.3 Estimateur non-biaisé

On dit qu'un estimateur est non biaisé si l'erreur d'estimation est de moyenne nulle :

$$E(\tilde{\theta}) = 0 \quad (12)$$

où $\tilde{\theta}$ est l'erreur d'estimation définie par :

$$\tilde{\theta} = \theta - \hat{\theta} \quad (13)$$

Un estimateur est non biaisé si 12 est vérifié pour tout k. On dit qu'il est asymptotiquement non biaisé si 12 est vérifié quand k tend vers l'infini.

1.4 Variance d'estimateur

Une deuxième caractéristique importante d'un estimateur est la variance de l'erreur d'estimation :

$$\mathrm{var}(\hat{\theta}(Z)) \triangleq E[(\hat{\theta}(Z) - E(\hat{\theta}(Z))^2] \quad (14)$$

Cette variance doit être aussi petite que possible, de façon à ce que l'estimé soit concentré autour de la vraie valeur du paramètre. La racine carré de la variance d'un estimateur[6] est l'écart type :

$$\sigma_{\theta} = \sqrt{\mathrm{var}(\hat{\theta})} \quad (15)$$

L'écart type fournit une mesure de précision pour l'estimateur.

[4] LS : Least Square

[5] MMSE : Minimum Mean Square Error

[6] MSE : Mean Square Error

1.5 Borne de Cramer Rao et Matrice d'Information de Fisher

Dans l'étude de problèmes d'estimation paramétrique, une inégalité très utile établit qu'une borne inférieure pour la variance de l'erreur d'estimés non biaisés existe. Cette limite est connue sous le nom de Borne de Cramer Rao[7].

Cas scalaire

Pour l'estimation d'un paramètre scalaire déterministe à partir d'un estimateur non biaisé, la variance est bornée par :

$$E[(\hat{\theta}(Z)-\theta_0)^2] \geq J^{-1} \tag{16}$$

où

$$J = -E\,[\frac{\partial^2 \ln p(Z/\theta)}{\partial \theta^2}]\bigg|_{\theta=\theta_0} = E\left\{[\frac{\partial \ln p(Z/\theta)}{\partial \theta}]^2\right\}\bigg|_{\theta=\theta_0} \tag{17}$$

est l'information de Fisher, $p(Z/\theta)$ est la fonction de vraisemblance, et θ_0 est la vraie valeur de θ.

Cas vectoriel

Pour un vecteur de paramètres déterministes, la matrice de covariance d'un estimateur non biaisé admet une borne minimale donnée par :

$$E[(\hat{\theta}(Z)-\theta_0)(\hat{\theta}(Z)-\theta_0)^t] \geq J^{-1} \tag{18}$$

où la matrice d'information de Fisher[8] est :

$$J \triangleq -E\,[\nabla_\theta \nabla_\theta^t \ln p(Z/\theta)]\big|_{\theta=\theta_0} = E\,[(\nabla_\theta \ln p(Z/\theta))(\nabla_\theta \ln p(Z/\theta))^t]\big|_{\theta=\theta_0} \tag{19}$$

où

$$\nabla_\theta = [\frac{\partial}{\partial \theta_1} ... \frac{\partial}{\partial \theta_n}]^t \tag{20}$$

est l'opérateur de gradient et n la dimension du vecteur de paramètres. La matrice d'information de Fisher peut être vue comme une quantification du (maximum) d'information existante pour un paramètre à partir d'observations.
Les estimateurs qui vérifient avec égalité la borne de Cramer Rao sont dits efficaces. La preuve de l'existence de la borne de Cramer Rao est donnée dans [2].

2 Estimation linéaire et non linéaire

2.1 Cas linéaire

Estimation des moindres carrées

Estimation batch

On souhaite estimer un vecteur θ de paramètres, de dimension n_θ, à partir d'un vecteur d'observations linéaires de dimension n_z :

$$z(i) = H(i)\theta + w(i), i = 1, ..., k \tag{21}$$

en minimisant l'erreur quadratique suivante :

$$J(k) = \sum_{i=1}^{k} [z(i) - H(i)\theta]^t R(i)^{-1} [z(i) - H(i)\theta] = [z^k - H^k\theta]^t (R^k)^{-1} [z^k - H^k\theta] \tag{22}$$

où

$$z^k = \begin{pmatrix} z(1) \\ \vdots \\ z(k) \end{pmatrix} \tag{23}$$

[7]CRLB : Cramer Rao Lower Bound
[8]FIM : Fisher Information Matrix

est le vecteur "empilé" de mesures (de dimension $kn_z \times 1$),

$$H^k = \begin{pmatrix} H(1) \\ \vdots \\ H(k) \end{pmatrix} \tag{24}$$

est la matrice "empilée" de mesure (de dimension $kn_z \times n_\theta$),

$$w^k = \begin{pmatrix} w(1) \\ \vdots \\ w(k) \end{pmatrix} \tag{25}$$

est le vecteur "empilé" d'erreurs de mesures, et

$$R^k = \begin{pmatrix} R(1) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & R(k) \end{pmatrix} = diag[R(i)] \tag{26}$$

est une matrice diagonale par bloc définie positive de dimension $kn_z \times kn_z$.
L'estimateur des moindres carrés qui minimise 22 est obtenu en annulant le gradient par rapport à θ :

$$\nabla_\theta J(k) = -2H^{k^t}(R^k)^{-1}[z^k - H^k\theta] = 0 \tag{27}$$

Finalement, l'estimateur est donné par :

$$\hat{\theta}(k) = [H^{k^t}(R^k)^{-1}H^k]^{-1}H^{k^t}(R^k)^{-1}z^k \tag{28}$$

De plus, la matrice de covariance de l'estimateur des moindres carrés est donnée par [2] :

$$P(k) = [H^{k^t}(R^k)^{-1}H^k]^{-1} \tag{29}$$

Estimation itérative

L'estimateur des moindres carrés peut être écrit d'une manière itérative. Dans ce cas, k est considéré comme un temps discret. L'obtention de $z(k+1)$ permet l'écriture des formes suivantes :

$$z^{k+1} = \begin{bmatrix} z^k \\ z(k+1) \end{bmatrix} \tag{30}$$

$$H^{k+1} = \begin{bmatrix} H^k \\ H(k+1) \end{bmatrix} \tag{31}$$

$$w^{k+1} = \begin{bmatrix} w^k \\ w(k+1) \end{bmatrix} \tag{32}$$

$$R^{k+1} = \begin{bmatrix} R^k & 0 \\ 0 & R(k+1) \end{bmatrix} \tag{33}$$

Les équations itératives sont données dans [2] :

$$\hat{\theta}(k+1) = \hat{\theta}(k) + W(k+1)[z(k+1) - H(k+1)\hat{\theta}(k)] \tag{34}$$

$$P(k+1) = P(k) - W(k+1)S(k+1)W(k+1)^t \tag{35}$$

où

$$W(k+1) \triangleq P(k)H(k+1)^tS(k+1)^{-1} \tag{36}$$

et

$$S(k+1) \triangleq H(k+1)P(k)H(k+1)^t + R(k+1) \tag{37}$$

Une initialisation est nécessaire puisque c'est une méthode itérative. Cela peut être réalisé en utilisant une technique dite "batch" sur un faible nombre de mesures ou en utilisant un estimé et une covariance associée à priori.

Ajustement polynomial On cherche à estimer les paramètres pour ajuster un polynôme de degré n à un ensemble d'observations. On considère que l'on observe, en présence de bruit, la distance r et la vitesse radiale $\dot{r}$ d'un véhicule dont l'évolution radiale (position,vitesse) est modélisée par des polynômes (fonction du temps). Le modèle de régression est définie par :

$$\begin{aligned} r(t) &= \sum_{j=0}^{n} a_j \frac{t^j}{j!} \\ \dot{r}(t) &= \sum_{j=1}^{n} a_j \frac{t^{j-1}}{(j-1)!} \end{aligned} \tag{38}$$

où les paramètres à estimer sont les coefficients a_j du polynôme. Ils correspondent au dérivée d'ordre j de la position au temps initial ($t = 0$). La méthode des moindres carrés est utilisée pour ajuster les paramètres du polynôme de degré n. Le modèle de régression est définie par 38, on a :

$$Z^k = H^k\theta + w^k \tag{39}$$

où w^k est donné par 25 et H^k par 24 avec :

$$H(i) = \begin{pmatrix} 1 & t_i & \dots & \dots & \frac{t_i^n}{n!} \\ 0 & 1 & t_i & \dots & \frac{t_i^{n-1}}{(n-1)!} \end{pmatrix} \tag{40}$$

On donne le vecteur estimé de paramètres par[1] :

$$\hat{\theta} = (H^{k^t} H^k)^{-1} H^{k^t} Z^k \tag{41}$$

et sa matrice de covariance par :

$$P(k) = \left(H^{k^t} R^{k^{-1}} H^k\right)^{-1} \tag{42}$$

où R^k est donnée par 26 avec :

$$R(i) = \begin{pmatrix} \sigma_r^2 & 0 \\ 0 & \sigma_{\dot{r}}^2 \end{pmatrix} \forall i \in [1...k] \tag{43}$$

On remarque ici que le vecteur de paramètres à estimer est indépendant des erreurs sur les observations. Finalement, après avoir déterminé les différents paramètres caractérisant la trajectoire de la cible, on peut prédire l'état de la cible pour un temps arbitraire. Si $x(t)$, au temps arbitraire t, est le vecteur d'état alors son estimé (prédiction) est :

$$\hat{x}(t/k) = F(t)\hat{\theta}(k) \tag{44}$$

où k est le nombre de mesures disponibles et $F(t)$ est la matrice définie par :

$$F(t) = \begin{pmatrix} 1 & t & \cdots & \frac{t^n}{n!} \\ 0 & 1 & \cdots & \frac{t^{n-1}}{(n-1)!} \end{pmatrix} \tag{45}$$

La covariance correspondante est donnée par :

$$P(t/k) = F(t)P(k)F(t)' \tag{46}$$

2.2 Cas non linéaire

On présente ici l'utilisation de la technique des moindres carrés pour estimer un vecteur θ de paramètres à partir de mesures non linéaires. Le vecteur "empilé" de mesures est donnée par :

$$z = z^k = h(\theta) + w^k \tag{47}$$

où h est une fonction non linéaire de θ (dimension n_θ).

Estimation itérative des moindres carrés : Méthode de Gauss-Newton Cette méthode, basée sur le principe des moindres carrés, est une technique qui améliore séquentiellement l'estimé courant en utilisant les mesures disponibles. A partir de l'estimé $\hat{\theta}_k$ à l'itération k, la mise à jour de l'estimé $\hat{\theta}_{k+1}$ est donnée par :

$$\hat{\theta}_{k+1} = \hat{\theta}_k + (J_k^t R^k J_k)^{-1} J_k^t R^{k^{-1}} [z^k - h(\hat{\theta}_k)] \tag{48}$$

où $J_k = \frac{\partial h}{\partial \theta}\big|_{\theta=\hat{\theta}_k}$ est la matrice jacobienne et R^k est la matrice de covariance définie par (26). La matrice jacobienne est donnée par :

$$J_k = \begin{pmatrix} \frac{\partial h(\theta(1))}{\partial \theta_1} & \cdots & \frac{\partial h(\theta(1))}{\partial \theta_{n_\theta}} \\ \vdots & & \vdots \\ \frac{\partial h(\theta(i))}{\partial \theta_1} & \cdots & \frac{\partial h(\theta(i))}{\partial \theta_{n_\theta}} \\ \vdots & & \vdots \\ \frac{\partial h(\theta(k))}{\partial \theta_1} & \cdots & \frac{\partial h(\theta(k))}{\partial \theta_{n_\theta}} \end{pmatrix}_{\theta=\hat{\theta}_k} \tag{49}$$

L'erreur moyenne au sens des moindres carrés de l'estimé $\hat{\theta}$ est :

$$E[(\hat{\theta}_k - \theta)(\hat{\theta}_k - \theta)^t] = (J^t R^{-1} J)^{-1} \tag{50}$$

Estimation du maximum de vraisemblance : Méthode de Newton-Raphson

L'estimé $\hat{\theta}^{MLE}$ est obtenu en maximisant la fonction de vraisemblance $\Lambda(\theta)$:

$$\Lambda(\theta) = \prod_{k=1}^{n} p(z(k)/\theta) \tag{51}$$

On a donc :

$$\hat{\theta}^{MLE} = \arg\max_{\theta} \Lambda(\theta) = \arg\min_{\theta} \{-\ln[\Lambda(\theta)]\} \triangleq \arg\min \lambda(\theta) \tag{52}$$

Cette minimisation est effectuée en utilisant la méthode de Newton-Raphson qui repose sur une approximation linéaire à l'ordre 1 de la fonction de vraisemblance. L'estimé $\hat{\theta}^k$; à l'instant k, est relié à l'estimé suivant $\hat{\theta}^{k+1}$ par :

$$\hat{\theta}_{k+1}^{MLE} = \hat{\theta}_k^{MLE} - Hess_k^{-1} \nabla_\theta \lambda(\theta)\big|_{\theta=\hat{\theta}_k^{MLE}} \tag{53}$$

où $Hess_k$ est la matrice Hessienne à l'instant k définie par :

$$Hess_k = \begin{pmatrix} \frac{\partial^2 \lambda(\theta)}{\partial \theta_1^2} & \cdots & \frac{\partial^2 \lambda(\theta)}{\partial \theta_1 \partial \theta_i} & \cdots & \frac{\partial^2 \lambda(\theta)}{\partial \theta_1 \partial \theta_{n_\theta}} \\ \vdots & \ddots & & & \vdots \\ \frac{\partial^2 \lambda(\theta)}{\partial \theta_i \partial \theta_1} & & \frac{\partial^2 \lambda(\theta)}{\partial \theta_i^2} & & \frac{\partial^2 \lambda(\theta)}{\partial \theta_i \partial \theta_{n_\theta}} \\ \vdots & & & \ddots & \vdots \\ \frac{\partial^2 \lambda(\theta)}{\partial \theta_{n_\theta} \partial \theta_1} & \cdots & \frac{\partial^2 \lambda(\theta)}{\partial \theta_{n_\theta} \partial \theta_i} & \cdots & \frac{\partial^2 \lambda(\theta)}{\partial \theta_{n_\theta}^2} \end{pmatrix}_{\theta=\hat{\theta}_k^{MLE}} \tag{54}$$

Références

[1] A. Antoniadis, J. Berruyer, and R. Carmona. *Régression non linéaire et applications.* Economica ISBN 2-7178-2344-1, 1992.

[2] Y. Bar-Shalom, X.R. Li, and T. Kirubarajan. Estimation with applications to tracking and navigation. In *New York : John Wiley and Sons.* 2001.