Quasi-Vraisemblance

Michel Goulard

Unité de Biométrie et d’Intelligence Artificielle, UR 875
Département de Mathématiques et Informatique Appliquées
INRA, Centre de Toulouse, Auzeville

Réunion réseau MSTGA 23 et 24 Nov. 2006

Méthodes d'estimation

$$X_1, \dots, X_n \text{ i.i.d. } f(x|\theta)$$

- moindres carrés : minimise $\sum (S(X_i) - g_i(\theta))^2$
- méthode des moments : résolution de $T_k(X) = M_k(\theta)$
- maximum de vraisemblance : maximise $\sum \log(f(X_i|\theta))$
- minimum contraste : minimum de $K_n(X|\theta)$ avec $K_n(X|\theta) \longrightarrow K(\theta, \theta_0)$

Introduction suite

- Quasi-vraisemblance introduite par Wedderburn pour famille exponentielle.
- McCullagh & Nelder dans GLM : corrélation, modèle aléatoire.
- C.C Heyde (1997) “Quasi-likelihood and its applications”, Springer.
- Godambe (1960), Durbin (1960), Godambe (1985), Heyde (1986), Godambe & Heyde (1987)
- Estimation de θ caract de la loi de X_t à partir de $\{X_t, t \in T\}$

1er Exemple Introductif

- $X_1, \ldots, X_T$ i.i.d. $E(X_t) = \theta$ $var(X_t) = \sigma^2$
- Gauss-Markov : moyenne empirique meilleur estimateur linéaire sans biais.
- $\mathcal{G} = \{G(X_1, \ldots, X_T) = \sum b_t(X_t - \theta), \sum b_t \neq 0\}$

$$G^{(s)} = (\sum b_t) \frac{\sum b_t(X_t - \theta)}{\sigma^2 \sum b_t^2}$$

G^* optimal dans $\mathcal{G}$ si $var(G^{*(s)}) \geq var(G^{(s)})$

- GM : $G^* = \sum(X_t - \theta)$ est optimal et $G^* = 0$ donne l'estimateur optimal.
- $G^{(s)} = -E(\nabla G)(E(G^2))^{-1}G$, $var(G^{(s)}) = -E(\nabla G^{(s)})$

2ème Exemple introductif

- $X_1, \ldots, X_T$ i.i.d. $E(X_t) = \alpha_t(\theta)$ $var(X_t) = \sigma_t^2(\theta)$

$$\mathcal{H} = \{H(\theta) = \sum b_t(\theta)(X_t - \alpha_t(\theta))\}$$

$$E(\nabla H) = \sum b_t(\theta)\nabla\alpha_t(\theta) \text{ et } E(H^2) = \sum b_t^2(\theta)\sigma_t^2(\theta)$$

- $H^{(s)} = (\sum b_t(\theta)\nabla\alpha_t(\theta))(\sum b_t^2(\theta)\sigma_t^2(\theta))^{-1}H$

$$H^* = \sum \frac{\nabla\alpha_t(\theta)}{\sigma_t^2(\theta)}(X_t - \alpha_t(\theta))$$

- Solution moindres carrés : solution de $H * + \sum(X_t - \alpha_t(\theta))^2 \nabla \frac{\sigma_t(\theta)}{\sigma_t^3(\theta)} = 0$

Lien avec vraisemblance

- $X_1, \dots, X_T$ i.i.d. $E(X_t) = \alpha_t(\theta)$, loi de X_t $f_t(X_t|\theta)$
- $\log L = \sum \log f_t(X_t|\theta)$

$$U = \frac{\partial \log L}{\partial \theta} \text{ et } H = \sum b_t(\theta)(X_t - \alpha_t(\theta))$$

- propriétés :$E(U) = 0$, $U^{(s)} = U$
  $E(UH) = -E(\nabla H)$, $corr^2(U,H) = var(H^{(s)})/E(U^2)$
- H^* est de corrélation max avec U
  $E((H^* - U)^2)$ est minimum
- si U est dans $\mathcal{H}$, U est optimal. U linéaire si famille exponentielle linéaire.

Définition générale

- $\{X_t, t \leq T\}$, $\{P_\theta, \theta \in I\!R^p\}$
- fonctions estimantes
  $\mathcal{G} = \{G_T(\{X_t, t \leq T\}, \theta)$ de dimension $p\}$
  $E(G_T(\theta)) = 0$, $E(\nabla G_T)$ et $E(G_T G_T')$ existent.
  On va considérer des sous-ensembles $\mathcal{H}$ de $\mathcal{G}$
- densité $p_T(\theta)$, $U_T(\theta) = p_T^{-1}(\theta)\nabla p_T(\theta)$
- standardisation $G_T^{(s)}(\theta) = -E(\nabla G_T)'(E(G_T G_T'))^{-1} G_T$
- $\mathcal{E}(G_T) = E(G_T^{(s)} G_T^{(s)'}) = E(\nabla G_T)'(E(G_T G_T'))^{-1} E(\nabla G_T)$

Définition générale (suite)

- G^* est optimal ou est un quasi-score, $G^* = 0$ donne l'estimateur de quasi-vraisemblance, G^* est O_F-optimal sur $\mathcal{H}$:

  $\mathcal{E}(G_T^*) - \mathcal{E}(G_T)$ est définie positive.

  $E((U_T - G_T^{(s)})(U_T - G_T^{(s)})') - E((U_T - G_T^{*(s)})(U_T - G_T^{*(s)})')$ est définie positive.

  $E((U_T - G_T^{*(s)})G_T^{(s)'}) = 0 = E(G_T^{(s)}(U_T - G_T^{*(s)})')$

Propriétés

- G_T^* est optimal si

$$E(G_T^{*(s)} G_T^{(s)'}) = E(G_T^{(s)} G_T^{*(s)'}) = E(G_T^{*(s)} G_T^{*(s)'})$$

équivalent à $E(\nabla G_T)^{-1} E(G_T G_T^*)$ constante sur $\mathcal{H}$

- $h_t(\theta)$ telle que $E(h_t(\theta)) = 0$ et $E(h_t(\theta) h_s(\theta)') = 0 \; s \neq t$ pour $H = \sum a_t(\theta) h_t(\theta)$ alors $a_t^* = E(\nabla h_t)'(E(h_t h_t'))^{-1}$
- Galton-Watson : $E(Z_1 | Z_0 = 1) = \theta$ et $E(Z_i | \mathcal{F}_{i-1}) = \theta Z_{i-1}$ $H^* = \sum \frac{1}{\sigma^2}(Z_i - \theta Z_{i-1})$ qui donne l'estimateur $(Z_1 + \ldots + Z_T)/(Z_0 + \ldots + Z_{T-1})$ qui est le maximum de vrais. pour une certaine famille de lois.

autres exemples

- On utilisera pour choisir des fonctions estimantes des scores obtenus pour certaines lois.
- $dX_t = \theta X_t dt + dW_t$

$$H = \int_0^T b_s dW_s = \int_0^T b_s (dX_s - \theta X_s ds)$$

$b_s^* = X_s$ et $\hat{\theta} = (\int_0^T X_s dX_s)/(\int_0^T X_s^2 ds)$

- Modèle Linéaire Généralisé : $Y = \mu(\theta) + e$, $E(e) = 0$, $cov(e) = V$
  $Q(\theta) = \nabla\mu' V^{-1}(Y - \mu(\theta))$
  c'est un quasi -score pour $\{A(\theta)(Y - \mu(\theta))\}$
- Les GEE sont les équations : Quasi-score =0

un autre exemple

- $X_i = \theta + e_i$
  $E(e_i|\mathcal{F}_{i-1}) = 0 \; E(e_i^2|\mathcal{F}_{i-1}) = \sigma^2 X_{i-1}^2$

$$H = \sum a_i(X_i - \theta) \text{ avec } a_i \; \mathcal{F}_{i-1}\text{-mesurable}$$

- Le quasi-score est $\frac{1}{\sigma^2}\sum \frac{(X_i-\theta)}{X_{i-1}^2}$.
- Wedderburn donne $\frac{1}{\sigma^2}\sum \frac{(X_i-\theta)}{E(X_{i-1}^2)}$.
  Le premier est meilleur au sens de la variance du quasi-score.

Généralisations pour martingale

- G_T une fonction estimante telle que
  $< G >_T = E(G_t G_t' | \mathcal{F}_{t-})$ existe et $< G >_T^{-1/2} G_T \longrightarrow N(0, I)$
  $\overline{G}_t(\theta) = \int_0^t E(\nabla dG_s(\theta) | \mathcal{F}_{s-})$
  dans le cas discret $E(G_t G_t' | \mathcal{F}_{t-1})$ et
  $\overline{G}_t(\theta) = \sum E(\nabla (G_s(\theta) - G_{s-1}(\theta)) | \mathcal{F}_{s-1})$
- G_T^* est O_A-optimal si : $\overline{G}_T^* < G^* >_T^{-1} \overline{G}_T^* - \overline{G}_T < G >_T^{-1} \overline{G}_T$ est p.s. définie positive.
- si $\overline{G}_T^{-1} < G, G^* >_T = \overline{G}_T^{*-1} < G^* >_T$ alors G_T^* est O_A-optimal
- si G_T^* est O_A-optimal et $\overline{G}_T^{*-1} < G^* >_T$ est non-aléatoire alors G^* est O_F-optimal.

Modèle signal + bruit

- $X_t = A_t + M_t$ et $M_t = \sum m_s$ ou $dX_t = dA_t + dM_s$
- fonctions estimantes
  $G_t = \sum \alpha_s(\theta) m_s(\theta)$ avec α_s $\mathcal{F}_{s-1}$-mesurable
  $G_t = \int_0^t \alpha_s(\theta) dM_s(\theta)$ avec α_s predictible
- solution de Hutton-Nelson :

$$G*_t = \sum (E(\nabla m_s|\mathcal{F}_{s-1})'(E(m_s m_s'|\mathcal{F}_{s-1})^- m_s$$

$$G*_t = \int_0^t (d\overline{M}_s)'(d < M >_s)^- dM_s$$

Deux exemples

- $dV_t = (-\rho V_t + \lambda)dt + dM(t)$ avec $< M >_t = \sigma^2 t$ (Kallianpur 1983) $\nabla dM(s) = -(-V_t, 1)$ et $d < M >_t = \sigma^2$
  $G_T^* = \frac{1}{\sigma^2} \int_0^T (-V_t, 1)'(dV_t - (-\rho V_t + \lambda)dt)$
  $$\begin{cases} \int_0^T V(t)dV(t) & = & \int_0^T (-\rho V_t + \lambda)V_t dt \\ \int_0^T dV(t) & = & \int_0^T (-\rho V_t + \lambda)dt \end{cases}$$
- $X_t = (\theta + \eta_t)X_{t-1} + \varepsilon_t$ avec
  $E(\eta_t|\mathcal{F}_{t-1}) = 0$ $E(\varepsilon_t|\mathcal{F}_{t-1}) = 0$ $E(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma^2 X_{t-1}$

  $X_t = \theta X_{t-1} + u_t$ $\nabla m_t = X_{t-1}$ et $G_T^* = \sum \frac{X_{t-1}}{E(u_t^2|\mathcal{F}_{t-1})}(X_t - \theta X_{t-1})$

  si η_t et ε_t indép. et $E(\eta_t^2|\mathcal{F}_{t-1}) = \tau^2$
  $E(u_t^2|\mathcal{F}_{t-1}) = X_{t-1}(\sigma^2 + \tau^2 X_{t-1})$ calcul possible.

Propriétés asymptotiques et tests

- Des développements asymptotiques sont faits : consistance, normalité pour les fonctions estimantes et pour les estimateurs.
- Zones de confiance : normalité et χ^2, optimalité des zones.
- Tests d'Hypothèses : $F'\theta = d$ contre $F'\theta \neq d$

Quasi-vraisemblance asymptotique

- préalables :
  A_n est asymptotiquement définie positive si il existe D_n telle que $A_n - D_n$ est définie positive et $||D_n|| \longrightarrow 0$

  $A = A^{1/2} A^{r/2}$

  $G_T^{(n)} = (E(\nabla G_T))^{-1} G_T$

- Q_T est un quasi-score asymptotique si

$$E(Q_T^{(n)} Q_T^{(n)\prime})^{-1/2} E(G_T^{(n)} G_T^{(n)\prime}) E(Q_T^{(n)} Q_T^{(n)\prime})^{-r/2} - I$$

  est asymptotiquement définie positive.

Modèle auto-normal

- $y_i | y_{[i]}$ suit une $N(\beta w_i' y, \tau_i^2)$

$$y \text{ suit une } N(0(I - \beta W)^{-1} M) \text{ avec } M = \mathrm{diag}(\tau_i^2)$$

- logvraisemblance donne le score

$$-y' W M^{-1} y + \frac{\partial log(|I - \beta W|)}{\partial \beta}$$

- approche quasi-vraisemblance : $\sum a_i (y_i - \beta w_i' y)$

$$G^* = \sum \frac{w_i' y (y_i - \beta w_i' y)}{\tau_i^2} = y' W M^{-1} y - \beta y' W M^{-1} W' y$$

- Pseudo-vraisemblance : maximum de $\sum \frac{1}{2\tau_i^2} (y_i - \beta w_i' y)^2$
  La nullité de la dérivée conduit à la même équation que la QV précédente.

Modèle conditionnels

Si $p(y_i|y_{[i]},\theta)$ on peut prendre $G = \sum a_i \frac{\partial \log p(y_i|y_{[i]},\theta)}{\partial\theta}$

$$G^* = \sum E(\frac{\partial^2 \log p_i(\theta)}{\partial\theta^2})'(E(\frac{\partial \log p_i(\theta)}{\partial\theta}\frac{\partial \log p_i(\theta)'}{\partial\theta}))^{-1}\frac{\partial \log p_i(\theta)}{\partial\theta}$$

Espérances conditionnelles ?
Relations avec PV ?
Intérêts ?

QV composée, Projection, Nuisance

- Plusieurs $G_{i,T}(\theta)$ on cherche $H = \sum \alpha_{i,T} G_{i,T}(\theta) = \alpha' G$

$$H^* \text{ à partir de } \alpha^{*'} = E(GG')^{-1} E(\nabla G)$$

- $F'\theta = d$, G fonct. estimante, $P_V = F(F'V^{-1}F)^- F'V^{-1}$

  $(M1)$ $G(\hat{\theta}) = 0$ et $\tilde{\theta} = (I - P_V)'\hat{\theta} + V^{-1}F(F'V^{-1}F)^- d$

  $(M2)$ $(I - P_V)G(\tilde{\theta}) = 0$ et $F'\tilde{\theta} = d$

  $(M3)$ $G(\tilde{\theta}) + F\lambda = 0$ et $F'\tilde{\theta} = d$
  Optimalité en prenant $V = \mathcal{E}(G)$, les méthodes sont équivalentes (au premier ordre), si Q est un Quasi-Score il y a optimalité par rapport à tous les G.
- $\theta' = (\phi', \psi')$ si Q est un Quasi-Score alors Q est optimal pour l'estimation de ϕ.

de EM à Projection-Solution

$X = (Y, Z)$, Y est observé

$$\frac{\partial \log L(\theta; Y)}{\partial \theta} = E_\theta\left(\frac{\partial \log L(\theta; X)}{\partial \theta} | Y\right)$$

étape maximisation : Max $E_{\theta_0}(\log L(\theta; X)|Y)$
moralement on cherche à résoudre $\frac{\partial \log L(\theta; Y)}{\partial \theta} = 0$ en résolvant

$$\frac{\partial}{\partial \theta} E_{\theta(p)}(\log L(\theta; X)|Y) = 0$$

à cause de

$$\frac{\partial \log L(\theta; Y)}{\partial \theta} = \left(\frac{\partial}{\partial \theta} E_{\theta_0}(\log L(\theta; X)|Y)\right)_{\theta_0 = \theta}$$

de EM à Projection-Solution

- $Q(\theta|X)$ Quasi-Score sur $\mathcal{H}_X$ pour X
- Projection :
  $H(\theta|\theta_0; Y)$ tel qu'il réalise le min pour G dans $\mathcal{H}_Y$ de

$$E_{\theta_0}(||G(\theta; Y) - Q(\theta|X)||^2)$$

- Solution : $\theta^{(p+1)}$ tel que

$$H(\theta^{(p+1)}|\theta^{(p)}; Y) = 0$$

- si $\theta^{(p)} \longrightarrow \tilde{\theta}$
  alors $H(\tilde{\theta}|\tilde{\theta}; Y)$ est un Quasi-Score sur $\mathcal{H}_Y$

Au-delà de la vraisemblance

- L'approche Quasi-Vraisemblance permet d'étendre le RELM au cas non-gaussien.
- Généralisation du score d'une diffusion à $dX_t = dA_t + dM_t$
- Champ de Markov caché :
  $y = \beta Wy + \varepsilon$ avec $E(\varepsilon_i|y_{[i]}) = 0$ $var(y_i|y_{[i]}) = \tau^2$
  observation de $z = By + \eta$, B connu, $E(\eta) = 0$,
  $var(\eta) = \sigma^2 I$

  Estimer y à β connu en combinant $A\varepsilon$ et $B\eta$
  $y = (\frac{I-\beta W}{\tau^2} + \frac{B'B}{\sigma^2})^{-1}\frac{B'z}{\sigma^2}$

  Estimer β avec $\varepsilon' S\varepsilon - E(\varepsilon' S\varepsilon)$, S matrice à trouver
  conduit à $y'Wy - tr(W(I-\beta W)^{-1}) = 0$
  (autre solution utiliser $y'Wy - \beta y'WW'y$)