Fonctions estimantes
dans les modèles linéaires généralisés
pour données corrélées

Jean-René Mathieu
Laboratoire de Statistique et Probabilités
Université Paul Sabatier-Toulouse

Journée INRIA -Grenoble- 8 Mars 2002

Mots-clés
- Quasi-score (quasi-vraisemblance)
- Modèle linéaire généralisé
- Fonctions estimantes
- Données répétées

1 Introduction

1.1 Exemple 1

Obs	group	init	rep
1	1	20	9
2	1	24	11
3	1	25	13
4	1	27	14
5	1	28	15
6	1	29	15
7	1	29	18
8	1	33	20
9	1	37	22
10	1	41	25
11	2	29	17
12	2	30	21
13	2	32	18
14	2	35	23
15	2	36	20
16	2	38	22
17	2	39	25
18	2	39	29
19	2	42	28
20	2	44	26

- "group" est un facteur à 2 niveaux
- "init" est une variable décrivant l'etat initial de l'individu
- "rep" est la variable réponse

Le but est d'étudier l'influence du facteur groupe sur la réponse en tenant compte de l'état initial ;
pour cela on "sait" que pour i=1,2 et j=1,2,...,10

- $E(Y_{i,j}) = \mu_{i,j} = \alpha_i + \beta_i.x_{i,j}$
- $Var(Y_{i,j}) = \phi.\mu_{i,j}^2$

1.2 Exemple 2

obs	group	init	y1	y2	y3
1	1	20	1	1	1
2	1	24	1	1	1
3	1	25	0	1	1
4	1	27	0	1	1
5	1	28	0	1	1
6	1	28	0	1	1
7	1	29	0	0	1
8	1	32	0	0	1
9	1	35	0	0	0
10	1	37	0	0	0
11	2	29	1	1	1
12	2	30	1	1	1
13	2	32	1	1	1
14	2	35	1	1	1
15	2	36	0	1	1
16	2	38	0	1	1
17	2	39	0	1	1
18	2	39	0	0	1
19	2	42	0	0	1
20	2	44	0	0	0

- "group" est un facteur à 2 niveaux
- " init" décrit l'état initial de l'individu
- la réponse est constituée de 3 observations dichotomiques dont on peut penser qu'elles sont corrélées

Le but de l'analyse est d'étudier l'influence de "group" sur la réponse en tenant compte de l'éventuelle influence de "init".
Pour i=1,2, j=1,2,...,10 et t=1,2,3, $Y_{i,j}^t$ désigne la réponse à l'instant t pour le j-iéme individu du goupe i.
On considère que $Y_{i,j}^t$ est une variable de Bernoulli de paramètre :

$$p_{i,j}^t = Pr[Y_{i,j}^t = 0] = \frac{exp(\alpha_i^t + \beta_i^t.xi,j)}{1 + exp(\alpha_i^t + \beta_i^t.xi,j)}$$

et que

$$Cor(Y_{i,j}^1, Y_{i,j}^2, Y_{i,j}^3) = C(\gamma)$$

Dans ces deux exemples :
- On ne sait pas écrire la vraisemblance $L(y, \theta)$
- Donc on ne sait pas écrire le score $\frac{DLogL(y,\theta)}{D\theta}$

1.3 Problème

On considère donc la situation suivante :
Pour $i = 1, 2, ..., n$, on observe une v.a. Y_i à valeur dans R^T :

$$Y_i = (Y_i^1, ..., Y_i^t, ..., Y_i^T)$$

avec

$$E(Y_i) = \mu_i = \mu(< x_i, \beta >), \beta \in R^p, \mu \text{ connue}$$

$$Var(Y_i^t) = \phi.V^t(\mu_i), \phi \in R^> \ , V^t \text{ connue}$$

$$Cor(Y_i) = C(\gamma), \gamma \in R^{p'}, C \text{ connue}$$

$$\text{donc } Var(Y_i) = \phi.V(\mu_i, \gamma) = \phi.V_i(\beta, \gamma), V_i \text{ matrice TxT connue}$$

β, γ, ϕ sont les paramètres inconnus

$$y_i = (y_i^1, ..., y_i^t, ..., y_i^T) \text{ est l'observation de } Y_i$$

$$y = (y_1, ..., y_i, ..., y_n) \text{ est l'observation de } Y = (Y_1, ..., Y_i, ..., Y_n)$$

2 Quasi-score

2.1 Définitions

- Quasi-score individuel

$$QS_i(Y_i, \beta, \gamma, \phi) = \frac{1}{\phi}(Y_i - \mu_i)V_i^{-1}(\beta, \gamma)\frac{D\mu_i}{D\beta}$$

- Quasi-score

$$QS(Y, \beta, \gamma, \phi) = \sum_{i=1}^{i=n} QS_i(Y_i, \beta, \gamma, \phi)$$

2.2 Propriétés

$$\forall(\beta, \gamma, \phi), E_{\beta,\gamma,\phi}(QS) = 0$$

$$\text{qui rappelle } E(\frac{DLogL(Y, \theta)}{D\theta}) = 0$$

$$\forall(\beta, \gamma, \phi), Var(QS) - E(\frac{DQS}{D\beta}) = 0$$

$$\text{qui rappelle } Var(\frac{DLogL(Y, \theta)}{D\theta}) - E(\frac{D^2 LogL(Y, \theta)}{D\theta^2}) = 0$$

2.3 Estimation de β et γ

$$QS(y,\hat{\beta},\hat{\gamma},\phi)=0 \Leftrightarrow \sum_{i=1}^{i=n}(y_i-\hat{\mu}_i).V_i^{-1}(\hat{\beta},\hat{\gamma}).\frac{D\mu_i}{D\beta}=0$$

Les estimations sont obtenues avec une procédure itérative qui peut être de la forme :

- étape initiale
  - $\beta^{(0)} \Rightarrow \mu_i^{(0)} \Rightarrow V^t(\mu_i^{(0)}) \Rightarrow rs_i^{t,(0)} = \frac{(y_i^t-\mu_i^{t,(0)})}{\sqrt{V^t(\mu_i^{(0)})}}$
  - $rs_i^{(0)} = (rs_i^{1,(0)}, rs_i^{2,(0)}, ..., rs_i^{T,(0)})$
  - $C_e^{(0)}$ est la matrice des corrélations empiriques des résidus $rs_i^{(0)}$
  - $\gamma^{(0)}$ est tel que $C(\gamma^{(0)})$ ajuste au mieux $C_e^{(0)}$
  - on a donc $V_i^{(0)}$
- première itération :
  - $\sum_{i=1}^{i=n}(y_i-\mu_i)(V_i^{(0)})^{-1}\frac{D\mu_i}{D\beta}=0 \Rightarrow \beta^{(1)} \Rightarrow \mu_i^{(1)} \Rightarrow V^t(\mu_i^{(1)}) \Rightarrow rs_i^{t,(1)}$
  - $rs_i^{(1)} \Rightarrow C_e^{(1)} \Rightarrow \gamma^{(1)} \Rightarrow V_i^{(1)}$
- etc...jusqu'à convergence (? ? ? ?)

2.4 Estimation de ϕ

$\hat{\phi} = \frac{1}{n.T-p}.sum_{t=1}^{t=T} sum_{i=1}^{i=n}(rs_i^t)^2$

2.5 Propriété

$$\mathcal{L}(\sqrt{n}(\hat{\beta}-\beta)) \underset{n\to\infty}{\longrightarrow} N_p[0,\phi.\Gamma_\gamma^{-1}(\beta)]$$

avec

$$\Gamma_\gamma(\beta) = \underset{n\to\infty}{lim}\frac{1}{n}\sum_{i=1}^{i=n}(\frac{D\mu_i}{D\beta})'V^{-1}(\mu_i,\gamma)\frac{D\mu_i}{D\beta} = \phi.\underset{n\to\infty}{lim}\frac{1}{n}\frac{DQS}{D\beta}$$

$\Gamma_\gamma(\beta)$ est "estimée" par $\frac{1}{n}\sum_{i=1}^{i=n}(\frac{D\mu_i}{D\beta})'_{\hat{\beta}}V_i^{-1}(\hat{\beta},\hat{\gamma})(\frac{D\mu_i}{D\beta})_{\hat{\beta}}$

2.6 Test d'une hypothèse H_0

$$H_0 \;: Q\beta = 0 \text{ où Q est qxp de rang q}$$

Test du quasi-score

$$T_1 = \frac{1}{n}||QS(y,\hat{\hat{\beta}},\hat{\gamma},\hat{\phi}||^2_{\hat{\phi}\Gamma^{-1}(\hat{\beta},\hat{\gamma})} \text{ où } \hat{\hat{\beta}} \text{ vérifie : } Q.\hat{\hat{\beta}} = 0$$

$$\text{Rejet de } H_0 \Leftrightarrow T_1 > \chi^2_{q,1-\alpha}$$

Test de Wald

$$T_2 = ||Q\hat{\beta}||^2_{(Var(Q\hat{\beta}))^{-1}}$$

$$\text{Rejet de } H_0 \Leftrightarrow T_2 > \chi^2_{q,1-\alpha}$$

3 Lien du quasi-score avec les fonctions estimantes

3.1 Définition

Une fonction estimante g d'un paramètre inconnu θ,$\theta \in R^m$, est $g(Y,\theta)$, à valeur dans R^m, telle que :

$$\forall\theta \in R^m, E_\theta[g(Y,\theta)] = 0$$

L'estimateur $\hat{\theta}$ par fonction estimante vérifie : $g(Y,\hat{\theta}) = 0$

Exemples d'estimateurs obtenus par fonction estimante

- estimateurs moindres carrés
- estimateurs max. de vraisemblance
- estimateurs par la méthode des moments
- estimateurs par χ^2 minimum
- estimateurs par quasi-score

3.2 Notion d'optimalité

Soit une classe G de fonctions estimantes pour un paramètre réel θ ; on dit que g^* est optimale ssi

$$\forall g \in G, \forall\theta \in R \frac{E(g^*)^2}{(E(\frac{Dg^*}{D\theta}))^2} \leq \frac{E(g)^2}{(E(\frac{Dg}{D\theta}))^2}$$

3.3 Propriété du quasi-score

Soit G la classe des fonctions estimantes

- linéaires en $Y - \mu(\beta)$, où $\mu(\beta)$ est la moyenne de Y : $g(Y - \mu(\beta))$ où g est une a.l. connue.
- telles que l'estimateur $\widetilde{\beta}$ par fonction estimante vérifie

$$\mathcal{L}(\sqrt{n}(\widetilde{\beta} - \beta)) \underset{n\to\infty}{\longrightarrow} N_p[0, \phi.\Delta(\beta)]$$

Alors

$$\forall\beta, \Delta(\beta) \geq \Gamma_\gamma^{-1}(\beta)$$

$$(\Delta(\beta) - \Gamma_\gamma^{-1}(\beta) \text{ est semi-définie positive})$$

Donc :

$$\forall a \in R^p, Var(< a, \hat{\beta} >) \leq Var(< a, \widetilde{\beta} >)$$

4 Lien avec le modèle linéaire généralisé

On suppose que $T = 1$; donc $V_i(\beta, \gamma) = V_i(\beta) = V(\mu_i)$

4.1 Définition

- Densité définie par

$$f(y, \mu, \phi) = 1_D.exp\{\frac{1}{\phi}[y.a(\mu) - b(\mu)] + c(y, \phi)\}, y \in R$$

$$E(Y) = \mu$$

$$Var(Y) = \phi.V^*(\mu) = \phi.\frac{1}{a'(\mu)}$$

  V^* est la fonction variance.
- $Y_1, Y_2, ..., Y_n$ sont indépendantes et Y_i est de densité $f(., \mu_i, \phi)$ avec $\mu_i = \mu(< x_i, \beta >)$ où μ^{-1} est la fonction de lien.
- Le log-vraisemblance du modèle saturé pour l'observation $y = (y_1, y_2, ..., y_n)$ est :

$$LogL(y, \mu, \phi) = \sum_{i=1}^{i=n} Logf(y_i, \mu_i, \phi)$$

- Le log-vraisemblance du modèle courant pour l'observation y est

$$LogL(y, \beta, \phi) = \sum_{i=1}^{i=n} Logf(y_i, \mu_i(\beta), \phi)$$

4.2 Estimation de β et ϕ par le maximum de vraisemblance

- score individuel
$$S_i(\beta) = (y_i - \mu_i).\frac{1}{\phi.V^*(\mu_i)}.\frac{D\mu_i}{D\beta}$$
- score
$$S(\beta) = \sum_{i=1}^{i=n}(y_i - \mu_i).\frac{1}{\phi.V^*(\mu_i)}.\frac{D\mu_i}{D\beta}$$
- $\hat{\beta}$ vérifie : $S(\hat{\beta}) = 0$
- $\hat{\phi}$ est obtenu par maximum de vraisemblance
- on definit alors

$$\text{Déviance} = 2\phi[LogL(y, y, \phi) - LogL(y, \hat{\beta}, \phi)]$$

qui sert à
  - mesurer l'adéquation du modéle aux observations si ϕ est connu
  - éventuellement estimer ϕ par $\frac{\text{Déviance}}{(n-p)}$ si ϕ est inconnu.

4.3 Propriété

Si la fonction variance V est aussi la fonction variance d'une distribution exponentielle $\widetilde{f}$, alors le quasi-score introduit en 2 et le score du modèle linéaire généralisé de densité $\widetilde{f}$ sont identiques

4.4 Log-Quasi-Vraisemblance et Quasi-Déviance

On définit.
- le log-quasi-densité de l'observation y_i :

$$LogQf(y_i, \mu_i, \phi) = \int_{y_i}^{\mu_i} \frac{y_i - u}{\phi.V(u)} du$$

- le log-quasi-vraisemblance du modèle saturé pour l'observation $y = (y_1, y_2, ..., y_n)$:

$$LogQL(y, \mu, \phi) = \sum_{i=1}^{i=n} LogQf(y_i, \mu_i, \phi)$$

- le log-quasi-vraisemblance du modèle courant pour l'observation $y = (y_1, y_2, ..., y_n)$:

$$LogQL(y, \beta, \phi) = \sum_{i=1}^{i=n} LogQf(y_i, \mu_i(\beta), \phi)$$

- et Quasi-déviance $= -2\phi.LogQL(y, \hat{\beta}, \phi)$

5 Traitement de l'exemple 1

```
rep                      =   group       init  init*group
 Response Distribution:  Quasi-Likelihood (Gamma Variance)
 Link Function:          Identity


                                  Nominal Variable Information
                                          Level  group
                                              1  1
                                              2  2


                                      Parameter Information
                                Parameter  Variable    group
                                        1  Intercept
                                        2  init*group  1
                                        3              2
                                        4  group       1
                                        5              2
                                        6  init


                                         Summary of Fit
Mean of Response  19.5500  Deviance       0.1184  Pearson ChiSq     0.1213
SCALE (Pearson)  131.8750  Deviance / DF  0.0074  Pearson ChiSq /DF 0.0076


                              Analysis of Deviance
             Source              DF    Deviance  Deviance / DF
            Model                 3     1.6445         0.5482
            Error                16     0.1184         0.0074
            C Total              19     1.7628          .

                                     Type III (Wald) Tests
                                                                 Pr >
                           Source             DF        ChiSq      ChiSq
                           init*group          1       0.6398      0.4238
                           group               1       1.1641      0.2806
                           init                1      96.9955      <.0001
```

Parameter Estimates

Variable	group	DF	Estimate	Std Error	ChiSq	Pr > ChiSq
Intercept		1	-1.7982	4.5789	0.1542	0.6945
init*group	1	1	0.7985	0.0756	111.6199	<.0001
	2	1	0.6785	0.1295	27.4383	<.0001
group	1	1	-5.3943	4.9995	1.1641	0.2806
	2	0	0.0000	.	.	.
init		0	0.0000	.	.	.

95% C.I. (Wald) for Parameters

Variable	group	Estimate	Lower	Upper
Intercept		-1.7982	-10.7727	7.1762
init*group	1	0.7985	0.6504	0.9466
	2	0.6785	0.4246	0.9324
group	1	-5.3943	-15.1932	4.4047
	2	0.0000	.	.
init		0.0000	.	

6 Traitement de l'exemple 2

```
proc genmod data=sasuser.gren2;
class ind group;
model rep=group init/dist=b type3 obstats;
repeated  subject=ind/type=exch corrw;
run;quit;
```

```
                              Model Information

                        Data Set                SASUSER.GREN2
                        Distribution                 Binomial
                        Link Function                   Logit
                        Dependent Variable                rep
                        Observations Used                  60

                       Criteria For Assessing Goodness Of Fit

        Criterion                 DF           Value        Value/DF

        Deviance                  57         47.6161          0.8354
        Scaled Deviance           57         47.6161          0.8354
        Pearson Chi-Square        57         44.1700          0.7749
        Scaled Pearson X2         57         44.1700          0.7749
        Log Likelihood                      -23.8081


           Algorithm converged.


                            GEE Model Information

                  Correlation Structure              Exchangeable
                  Subject Effect                  ind (20 levels)
                  Number of Clusters                           20
                  Correlation Matrix Dimension                  3
                  Maximum Cluster Size                          3
                  Minimum Cluster Size                          3


           Algorithm converged.
```

Working Correlation Matrix

	Col1	Col2	Col3
Row1	1.0000	-0.3326	-0.3326
Row2	-0.3326	1.0000	-0.3326
Row3	-0.3326	-0.3326	1.0000

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

Parameter		Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept		-19.6017	2.9961	-25.4740	-13.7294	-6.54	<.0001
group	1	5.2451	0.8765	3.5271	6.9631	5.98	<.0001
group	2	0.0000	0.0000	0.0000	0.0000	.	.
init		0.4998	0.0773	0.3483	0.6514	6.47	<.0001

Score Statistics For Type 3 GEE Analysis

Source	DF	Chi-Square	Pr > ChiSq
group	1	9.40	0.0022
init	1	9.34	0.0022

Remarque 1 Dans la procedure Genmod de SAS, on peut ,pour la distribution de Bernoulli, remplacer la modélisation de $C(\gamma)$ par la modélidation du "odd-ratio"

Remarque 2 Toujours à propos de réponses qualitatives répétées une alternative aux methodes utilisées ci-dessus consiste à modeliser les transitions par des chaines de Markov dans lesquelles les probabilités de transition s'expriment en fonction de variables exogènes selon des modéles linéaires généralisés.

Une question Comment faire du ré-échantillonage dans les fonctions estimantes ?
Une réponse est proposée dans un chapitre du “Godambe” :

- Si les v.a.r. Z_i sont de moyenne μ et de variance σ_i^2
- Si les v.a.r. T_i sont indépendantes, centrées et réduites

alors les v.a.r.

$$U_i = \bar{Z} + (Z_i - \bar{Z})T_i$$

peuvent jouer le rôle d'un ré-échantillonage des Z_i en ce sens que

$$||F_n - G_n|| \underset{n\to\infty}{\longrightarrow} 0$$

- F_n est la fonction de répartition de $\sqrt{n}(\bar{Z} - \mu)$
- G_n est la fonction de répartition $\sqrt{n}(\bar{Z} - \bar{U})$

On applique ceci à $Z_i = QS_i$ avec $\mu = 0$ (donc $\bar{Z} = 0$)

Bibliographie

- P. McCullagh et J.A. Nelder : *Generalised linear models*- Chapman et Hall - Londres, New-york.1989
- V.P. Godambe : *Estimating Functions* -Oxford Science Publications.1991 Clarended Press- Oxford
- Notice de SAS (procédure Genmod)

Bibliographie additionnelle sur les méthodes re ré-echantillonage en régréssion

- B.EFRON : Bootstrap methods : another look at the Jackknife, *The Annals of Statistics* Vol. 7 ,pp1-26,1979
- B.EFRON : The Jackknife, the bootstrap andd other resampling plans Siam, Philadelphia, 1982.
- B.EFRON : Bootstrap confidence intervals for a class of parametric problems, *Biometrika*, vol. 72,pp 42-58, 1985.
- B.EFRON : Better bootstrap confidence intervals, *The Journal of the American Statistical Association*,vol. 82, $n^0$397,pp 171-200,1987
- D.A. FREEDMAN : Bootstapping regression models, *The Annals of Statistics* Vol. 9,pp 1218-1228,1981
- P. HALL : On the bootstrap and confidence intervals, *The Annals of Statistics* Vol. 14,pp 1431-1452,1981
- D.V. HINKLEY : Bootstrap methods, *The Journal of the Royal Statistical Society* ,B,vol. 50,$n^0$3,pp. 321-337,1988
- L.H. MOULTON,S.L. ZEGER :Analysing repeated measures on generalized linear models via the bootstrap, *Biometrics*, vol. 45,,pp. 381-394,1989
- L.H. MOULTON,S.L. ZEGER :Bootstraping generalized linear models, *Computational Statistics and Data Analysis*vol. 11,pp 53-63,1991
- J. SHAO : Bootstrap Variance and bias estimation in linear models, *The Canadian Journal of Statistics* vol. 16, pp 371-382,1988
- J. SHAO : Ordinary and weighted least-square estimators, *The Canadian Journal of Statistics* vol. 18, pp 327-336,1990
- J. SHAO,D. TU : *The Jackknife and Bootstrap* Editions Springer 1995
- R.A. STINE : Bootstrap prediction intervals for regression, *The Journal of the American Statistical Association*,vol. 80,$n^0$392,pp 1026- 1031,1985
- C.F.J. WU : Jackknife,bootstrap and other resampling methods in regression analysis, *The Annals of Statistics* , vol. 14,pp 1261-1350 ,1986