Quelques modèles logistiques polytomiques

L. Rouvière
laurent.rouviere@univ-rennes2.fr

JANVIER 2015

1 Introduction

2 Le modèle saturé

3 Le modèle de régression logistique multinomial

4 Le modèle de régression logistique ordinal

Bibliographie

1 Introduction

2 Le modèle saturé

3 Le modèle de régression logistique multinomial

4 Le modèle de régression logistique ordinal
Bibliographie

Introduction

Modèles polytomiques

- Jusqu'à présent, nous étions dans un contexte où la variable à expliquer était binaire à valeurs dans $\{0, 1\}$.
- Dans deux nombreux cas, on peut être amener à expliquer une variable qualitative prenant plus de deux modalités (scrutin à plus de deux candidats, degrés de satisfaction pour un produit, mention à un examen...)

Objectif

Etendre le modèle de régression logistique (binaire) à un cadre polytomique.

Modèles polytomiques

- Jusqu'à présent, nous étions dans un contexte où la variable à expliquer était binaire à valeurs dans $\{0, 1\}$.
- Dans deux nombreux cas, on peut être amener à expliquer une variable qualitative prenant plus de deux modalités (scrutin à plus de deux candidats, degrés de satisfaction pour un produit, mention à un examen...)

Objectif

Etendre le modèle de régression logistique (binaire) à un cadre polytomique.

Exemple

On pose à 195 étudiants la question : si vous trouvez un portefeuille dans la rue contenant de l'argent et des papiers :

- vous gardez tout (réponse 1) ;
- vous gardez l'argent et rendez le portefeuille (réponse 2) ;
- vous rendez tout (réponse 3).

On construit alors la variable `WALLET` telle que

- `WALLET=1` si l'étudiant répond 1 ;
- `WALLET=2` si l'étudiant répond 2 ;
- `WALLET=3` si l'étudiant répond 3 ;

Exemple

On pose à 195 étudiants la question : si vous trouvez un portefeuille dans la rue contenant de l'argent et des papiers :

- vous gardez tout (réponse 1) ;
- vous gardez l'argent et rendez le portefeuille (réponse 2) ;
- vous rendez tout (réponse 3).

On construit alors la variable `WALLET` telle que

- `WALLET`=1 si l'étudiant répond 1 ;
- `WALLET`=2 si l'étudiant répond 2 ;
- `WALLET`=3 si l'étudiant répond 3 ;

Pour chaque étudiant, on note :

- Le sexe (variable `MALE`=1 si homme, 0 si femme) ;
- la nature des études suivies (variable `BUSINESS`= 1 pour les écoles de commerce, 0 pour les autres écoles) ;
- l'existence de punitions passées (variable `PUNISH`=1 si puni seulement à l'école primaire, 2 si puni seulement à l'école primaire et secondaire et 3 si puni seulement à l'école primaire, secondaire et supérieur) ;
- l'explication ou pas par les parents des punitions reçues dans l'enfance (variable `EXPLAIN`=1 si les parents expliquaient, 0 sinon).

On cherche à expliquer la variable `WALLET` par les autres variables.

Pour chaque étudiant, on note :

- Le sexe (variable `MALE`=1 si homme, 0 si femme) ;
- la nature des études suivies (variable `BUSINESS`= 1 pour les écoles de commerce, 0 pour les autres écoles) ;
- l'existence de punitions passées (variable `PUNISH`=1 si puni seulement à l'école primaire, 2 si puni seulement à l'école primaire et secondaire et 3 si puni seulement à l'école primaire, secondaire et supérieur) ;
- l'explication ou pas par les parents des punitions reçues dans l'enfance (variable `EXPLAIN`=1 si les parents expliquaient, 0 sinon).

On cherche à expliquer la variable `WALLET` par les autres variables.

Les données

- Les données brûtes sont présentées sous forme individuelles ($n = 195$) :

```
> donnees[1:5,]
  wallet male business punish explain
1      2    0        0      2       0
2      2    0        0      2       1
3      3    0        0      1       1
4      3    0        0      2       0
5      1    1        0      1       1
```

- Il y a cependant des répétitions... Et pour que certains indicateurs soient bien calculés (comme la déviance), il faut les présenter sous forde de données répétées ($T = 23$) :

```
> donnees1[1:5,]
  male business punish explain wallet.A A wallet.B B wallet.C  C
1    0        0      1       0        1 1        2 3        3  8
2    0        0      1       1        1 0        2 5        3 45
3    0        0      2       0        1 0        2 2        3  5
4    0        0      2       1        1 0        2 2        3  5
5    0        0      3       0        1 3        2 1        3  1
```

Les données

- Les données brûtes sont présentées sous forme individuelles ($n = 195$) :

```
> donnees[1:5,]
  wallet male business punish explain
1      2    0        0      2       0
2      2    0        0      2       1
3      3    0        0      1       1
4      3    0        0      2       0
5      1    1        0      1       1
```

- Il y a cependant des répétitions... Et pour que certains indicateurs soient bien calculés (comme la déviance), il faut les présenter sous forde de données répétées ($T = 23$) :

```
> donnees1[1:5,]
  male business punish explain wallet.A A wallet.B B wallet.C  C
1    0        0      1       0        1 1        2 3        3  8
2    0        0      1       1        1 0        2 5        3 45
3    0        0      2       0        1 0        2 2        3  5
4    0        0      2       1        1 0        2 2        3  5
5    0        0      3       0        1 3        2 1        3  1
```

Notations

- On cherche à expliquer une variable Y à K modalités $\{1,\dots,K\}$ par p variables explicatives $X_1,\dots,X_p$.
- Là encore, il convient de distinguer les données individuelles des données répétées.

- Données individuelles : il n'y a pas de répétitions sur les x_i. Les données sont $\{(x_1,y_1),\dots,(x_n,y_n), x_i\in\mathbb{R}^p, y_i\in\{1,\dots,K\}\}$.
- Données répétées : $\{(x_t,n_t,s_{1t},\dots,s_{Kt}), t=1,\dots,T\}$ où
  - n_t : nombre de répétitions au point x_t.
  - s_{jt} : nombre de fois où la modalité j de Y a été observée au point x_t :

$$s_{jt}=\sum_{i=1}^{n_t}\mathbf{1}_{y_{it}=j},\quad j=1,\dots,K,\quad t=1,\dots,T.$$

Notations

- On cherche à expliquer une variable Y à K modalités $\{1, \ldots, K\}$ par p variables explicatives $X_1, \ldots, X_p$.
- Là encore, il convient de distinguer les données individuelles des données répétées.

- Données individuelles : il n'y a pas de répétitions sur les x_i. Les données sont $\{(x_1, y_1), \ldots, (x_n, y_n), x_i \in \mathbb{R}^p, y_i \in \{1, \ldots, K\}\}$.
- Données répétées : $\{(x_t, n_t, s_{1t}, \ldots, s_{Kt}), t = 1, \ldots, T\}$ où
  - n_t : nombre de répétitions au point x_t.
  - s_{jt} : nombre de fois où la modalité j de Y a été observée au point x_t :

$$s_{jt} = \sum_{i=1}^{n_t} \mathbf{1}_{y_{it}=j}, \quad j = 1, \ldots, K, \quad t = 1, \ldots, T.$$

Notations

- On cherche à expliquer une variable Y à K modalités $\{1, \dots, K\}$ par p variables explicatives $X_1, \dots, X_p$.
- Là encore, il convient de distinguer les données individuelles des données répétées.

- Données individuelles : il n'y a pas de répétitions sur les x_i. Les données sont $\{(x_1, y_1), \dots, (x_n, y_n), x_i \in \mathbb{R}^p, y_i \in \{1, \dots, K\}\}$.
- Données répétées : $\{(x_t, n_t, s_{1t}, \dots, s_{Kt}), t = 1, \dots, T\}$ où
  - n_t : nombre de répétitions au point x_t.
  - s_{jt} : nombre de fois où la modalité j de Y a été observée au point x_t :

$$s_{jt} = \sum_{i=1}^{n_t} \mathbf{1}_{y_{it}=j}, \quad j = 1, \dots, K, \quad t = 1, \dots, T.$$

Notations

- On cherche à expliquer une variable Y à K modalités $\{1, \dots, K\}$ par p variables explicatives $X_1, \dots, X_p$.
- Là encore, il convient de distinguer les données individuelles des données répétées.

- Données individuelles : il n'y a pas de répétitions sur les x_i. Les données sont $\{(x_1, y_1), \dots, (x_n, y_n), x_i \in \mathbb{R}^p, y_i \in \{1, \dots, K\}\}$.
- Données répétées : $\{(x_t, n_t, s_{1t}, \dots, s_{kt}), t = 1, \dots, T\}$ où
  - n_t : nombre de répétitions au point x_t.
  - s_{jt} : nombre de fois où la modalité j de Y a été observée au point x_t :

$$s_{jt} = \sum_{i=1}^{n_t} \mathbf{1}_{y_{it}=j}, \quad j = 1, \dots, K, \quad t = 1, \dots, T.$$

Début de modélisation

- On se place dans le cas de données répétées : les s_{jt} sont des réalisations de variables aléatoires S_{jt}.
- Si les observations sont indépendantes alors le vecteur réponse $(S_{1t}, \ldots, S_{Kt})$ suit une loi multinomiale de paramètres $(n_t, p_1(x_t), \ldots, p_K(x_t))$ avec

$$p_j(x_t) = \mathbf{P}(Y_t = j).$$

Poser un modèle revient à spécifier des contraintes de forme entre les probabilités $p_j(x_t)$.

Début de modélisation

- On se place dans le cas de données répétées : les s_{jt} sont des réalisations de variables aléatoires S_{jt}.
- Si les observations sont indépendantes alors le vecteur réponse $(S_{1t}, \dots, S_{Kt})$ suit une loi multinomiale de paramètres $(n_t, p_1(x_t), \dots, p_K(x_t))$ avec

$$p_j(x_t) = \mathbf{P}(Y_t = j).$$

Poser un modèle revient à spécifier des contraintes de forme entre les probabilités $p_j(x_t)$.

Début de modélisation

- On se place dans le cas de données répétées : les s_{jt} sont des réalisations de variables aléatoires S_{jt}.
- Si les observations sont indépendantes alors le vecteur réponse $(S_{1t}, \dots, S_{Kt})$ suit une loi multinomiale de paramètres $(n_t, p_1(x_t), \dots, p_K(x_t))$ avec

$$p_j(x_t) = \mathbf{P}(Y_t = j).$$

Poser un modèle revient à spécifier des contraintes de forme entre les probabilités $p_j(x_t)$.

Les modèles étudiés

- Il y a plein de façons de poser des contraintes entre les probabilités $p_j(x_t)$.
- On peut donc définir un grand nombre de modèles.
- Nous présontons dans ce chapitre les modèles les plus classiques :

1. le modèle saturé ;
2. Le modèle de régression logistique multinomial ;
3. Le modèle logistique ordinal (ou modèle à égalité des pentes), valable dans le cas où Y est une variable ordinale.

Les modèles étudiés

- Il y a plein de façons de poser des contraintes entre les probabilités $p_j(x_t)$.
- On peut donc définir un grand nombre de modèles.
- Nous présontons dans ce chapitre les modèles les plus classiques :

1. le modèle saturé ;
2. Le modèle de régression logistique multinomial ;
3. Le modèle logistique ordinal (ou modèle à égalité des pentes), valable dans le cas où Y est une variable ordinale.

Le modèle saturé

Présentation

- Tout comme dans le cas binaire, ce modèle ne pose pas de restrictions entre les probabilités $p_j(x_t)$.
- Il contient donc comme paramètres les probabilités

$$p_j(x_t) = \mathbf{P}(Y_t = j), \quad j = 1, \dots, K, \quad t = 1, \dots, T.$$

- Comme $\sum_{j=1}^{K} p_j(x_t) = 1$, ce modèle comprend $T(K-1)$ paramètres inconnus.

EMV

Il est facile de voir que les estimateurs du maximum de vraisemblance sont donnés par $\hat{p}_j(x_t) = S_{jt}/n_t$.

Présentation

- Tout comme dans le cas binaire, ce modèle ne pose pas de restrictions entre les probabilités $p_j(x_t)$.
- Il contient donc comme paramètres les probabilités

$$p_j(x_t) = \mathbf{P}(Y_t = j), \quad j = 1, \dots, K, \quad t = 1, \dots, T.$$

- Comme $\sum_{j=1}^{K} p_j(x_t) = 1$, ce modèle comprend $T(K-1)$ paramètres inconnus.

EMV

Il est facile de voir que les estimateurs du maximum de vraisemblance sont donnés par $\hat{p}_j(x_t) = S_{jt}/n_t$.

Présentation

- Tout comme dans le cas binaire, ce modèle ne pose pas de restrictions entre les probabilités $p_j(x_t)$.
- Il contient donc comme paramètres les probabilités

$$p_j(x_t) = \mathbf{P}(Y_t = j), \quad j = 1, \dots, K, \quad t = 1, \dots, T.$$

- Comme $\sum_{j=1}^{K} p_j(x_t) = 1$, ce modèle comprend $T(K-1)$ paramètres inconnus.

EMV

Il est facile de voir que les estimateurs du maximum de vraisemblance sont donnés par $\hat{p}_j(x_t) = S_{jt}/n_t$.

Présentation

- Tout comme dans le cas binaire, ce modèle ne pose pas de restrictions entre les probabilités $p_j(x_t)$.
- Il contient donc comme paramètres les probabilités

$$p_j(x_t) = \mathbf{P}(Y_t = j), \quad j = 1, \dots, K, \quad t = 1, \dots, T.$$

- Comme $\sum_{j=1}^K p_j(x_t) = 1$, ce modèle comprend $T(K-1)$ paramètres inconnus.

EMV

Il est facile de voir que les estimateurs du maximum de vraisemblance sont donnés par $\hat{p}_j(x_t) = S_{jt}/n_t$.

Propriétés

- On note $\Pi_t = (p_1(x_t), \ldots, p_K(x_t))$ et $\hat{\Pi}_t = (\hat{p}_1(x_t), \ldots, \hat{p}_K(x_t))$.

Proposition

On a

1. $\hat{\Pi}_t$ est un estimateur sans biais de Π_t.
2. $\mathbf{V}(\hat{\Pi}_t) = \frac{1}{n_t}\Sigma_t$ où $\Sigma_t(j,j) = p_j(x_t)(1 - p_j(x_t))$ et $\Sigma_t(j,\ell) = -p_j(x_t)p_\ell(x_t)$ si $1 \leq j \neq \ell \leq K$.
3. Si $n_t \to \infty$ alors
$$\sqrt{n_t}(\hat{\Pi}_t - \Pi_t) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma_t).$$

Inconvénients

- On doit disposer d'un grand nombre de répétitions à chaque point du design pour que les estimateurs soient précis.
- Ce modèle ne nous renseigne pas sur les probabilités $\mathbf{P}(Y = j)$ pour des points x n'appartenant pas au design.

Propriétés

- On note $\Pi_t = (p_1(x_t), \dots, p_K(x_t))$ et $\hat{\Pi}_t = (\hat{p}_1(x_t), \dots, \hat{p}_K(x_t))$.

Proposition

On a

1. $\hat{\Pi}_t$ est un estimateur sans biais de Π_t.
2. $\mathbf{V}(\hat{\Pi}_t) = \frac{1}{n_t}\Sigma_t$ où $\Sigma_t(j,j) = p_j(x_t)(1 - p_j(x_t))$ et $\Sigma_t(j,\ell) = -p_j(x_t)p_\ell(x_t)$ si $1 \leq j \neq \ell \leq K$.
3. Si $n_t \to \infty$ alors
$$\sqrt{n_t}(\hat{\Pi}_t - \Pi_t) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma_t).$$

Inconvénients

- On doit disposer d'un grand nombre de répétitions à chaque point du design pour que les estimateurs soient précis.
- Ce modèle ne nous renseigne pas sur les probabilités $\mathbf{P}(Y = j)$ pour des points x n'appartenant pas au design.

Propriétés

- On note $\Pi_t = (p_1(x_t), \dots, p_K(x_t))$ et $\hat{\Pi}_t = (\hat{p}_1(x_t), \dots, \hat{p}_K(x_t))$.

Proposition

On a

1. $\hat{\Pi}_t$ est un estimateur sans biais de Π_t.
2. $\mathbf{V}(\hat{\Pi}_t) = \frac{1}{n_t}\Sigma_t$ où $\Sigma_t(j,j) = p_j(x_t)(1 - p_j(x_t))$ et $\Sigma_t(j,\ell) = -p_j(x_t)p_\ell(x_t)$ si $1 \leq j \neq \ell \leq K$.
3. Si $n_t \to \infty$ alors
$$\sqrt{n_t}(\hat{\Pi}_t - \Pi_t) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma_t).$$

Inconvénients

- On doit disposer d'un grand nombre de répétitions à chaque point du design pour que les estimateurs soient précis.
- Ce modèle ne nous renseigne pas sur les probabilités $\mathbf{P}(Y = j)$ pour des points x n'appartenant pas au design.

Propriétés

- On note $\Pi_t = (p_1(x_t), \dots, p_K(x_t))$ et $\hat{\Pi}_t = (\hat{p}_1(x_t), \dots, \hat{p}_K(x_t))$.

Proposition

On a

1. $\hat{\Pi}_t$ est un estimateur sans biais de Π_t.
2. $\mathbf{V}(\hat{\Pi}_t) = \frac{1}{n_t}\Sigma_t$ où $\Sigma_t(j,j) = p_j(x_t)(1 - p_j(x_t))$ et $\Sigma_t(j,\ell) = -p_j(x_t)p_\ell(x_t)$ si $1 \leq j \neq \ell \leq K$.
3. Si $n_t \to \infty$ alors
$$\sqrt{n_t}(\hat{\Pi}_t - \Pi_t) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma_t).$$

Inconvénients

- On doit disposer d'un grand nombre de répétitions à chaque point du design pour que les estimateurs soient précis.
- Ce modèle ne nous renseigne pas sur les probabilités $\mathbf{P}(Y = j)$ pour des points x n'appartenant pas au design.

Propriétés

- On note $\Pi_t = (p_1(x_t), \dots, p_K(x_t))$ et $\hat{\Pi}_t = (\hat{p}_1(x_t), \dots, \hat{p}_K(x_t))$.

Proposition

On a

1. $\hat{\Pi}_t$ est un estimateur sans biais de Π_t.
2. $\mathbf{V}(\hat{\Pi}_t) = \frac{1}{n_t}\Sigma_t$ où $\Sigma_t(j,j) = p_j(x_t)(1 - p_j(x_t))$ et $\Sigma_t(j,\ell) = -p_j(x_t)p_\ell(x_t)$ si $1 \leq j \neq \ell \leq K$.
3. Si $n_t \to \infty$ alors
$$\sqrt{n_t}(\hat{\Pi}_t - \Pi_t) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma_t).$$

Inconvénients

- On doit disposer d'un grand nombre de répétitions à chaque point du design pour que les estimateurs soient précis.
- Ce modèle ne nous renseigne pas sur les probabilités $\mathbf{P}(Y = j)$ pour des points x n'appartenant pas au design.

Propriétés

- On note $\Pi_t = (p_1(x_t), \dots, p_K(x_t))$ et $\hat{\Pi}_t = (\hat{p}_1(x_t), \dots, \hat{p}_K(x_t))$.

Proposition

On a

1. $\hat{\Pi}_t$ est un estimateur sans biais de Π_t.
2. $\mathbf{V}(\hat{\Pi}_t) = \frac{1}{n_t}\Sigma_t$ où $\Sigma_t(j,j) = p_j(x_t)(1 - p_j(x_t))$ et $\Sigma_t(j,\ell) = -p_j(x_t)p_\ell(x_t)$ si $1 \leq j \neq \ell \leq K$.
3. Si $n_t \to \infty$ alors
$$\sqrt{n_t}(\hat{\Pi}_t - \Pi_t) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma_t).$$

Inconvénients

- On doit disposer d'un grand nombre de répétitions à chaque point du design pour que les estimateurs soient précis.
- Ce modèle ne nous renseigne pas sur les probabilités $\mathbf{P}(Y = j)$ pour des points x n'appartenant pas au design.

Le modèle de régression logistique multinomial

- On dispose d'une variable explicative Y à K modalités et on cherche à modéliser les probabilités

$$\mathbf{P}(Y_t = j), \quad j = 1, \dots, K-1, \quad t = 1, \dots, T.$$

- L'approche consiste à se donner une modalité de référence, par exemple la modalité K, et à modéliser les prbabilités $p_j(x)$ selon

$$\log \frac{p_j(x_t)}{p_K(x_t)} = \beta_{1j}x_{t1} + \dots + \beta_{pj}x_{tp} = x_t'\beta_j,$$

où $\beta_j = (\beta_{1j}, \dots, \beta_{pj})$.

Remarques

- Le modèle ne dépend pas de la modalité de référence choisie ! (seule la valeur des coefficients, et donc leur interprétation, en dépend).
- Ce modèle comprend $p(K-1)$ paramètres à estimer.
- Si $K = 2$, on retombe sur le modèle logistique binaire.

- On dispose d'une variable explicative Y à K modalités et on cherche à modéliser les probabilités

$$\mathbf{P}(Y_t = j), \quad j = 1, \ldots, K-1, \quad t = 1, \ldots, T.$$

- L'approche consiste à se donner une modalité de référence, par exemple la modalité K, et à modéliser les prbabilités $p_j(x)$ selon

$$\log \frac{p_j(x_t)}{p_K(x_t)} = \beta_{1j} x_{t1} + \ldots + \beta_{pj} x_{tp} = x_t' \beta_j,$$

où $\beta_j = (\beta_{1j}, \ldots, \beta_{pj})$.

Remarques

- Le modèle ne dépend pas de la modalité de référence choisie ! (seule la valeur des coefficients, et donc leur interprétation, en dépend).
- Ce modèle comprend $p(K-1)$ paramètres à estimer.
- Si $K = 2$, on retombe sur le modèle logistique binaire.

- On dispose d'une variable explicative Y à K modalités et on cherche à modéliser les probabilités

$$\mathbf{P}(Y_t = j), \quad j = 1, \dots, K-1, \quad t = 1, \dots, T.$$

- L'approche consiste à se donner une modalité de référence, par exemple la modalité K, et à modéliser les prbabilités $p_j(x)$ selon

$$\log \frac{p_j(x_t)}{p_K(x_t)} = \beta_{1j} x_{t1} + \ldots + \beta_{pj} x_{tp} = x_t' \beta_j,$$

où $\beta_j = (\beta_{1j}, \dots, \beta_{pj})$.

Remarques

- Le modèle ne dépend pas de la modalité de référence choisie ! (seule la valeur des coefficients, et donc leur interprétation, en dépend).
- Ce modèle comprend $p(K-1)$ paramètres à estimer.
- Si $K = 2$, on retombe sur le modèle logistique binaire.

- On dispose d'une variable explicative Y à K modalités et on cherche à modéliser les probabilités

$$\mathbf{P}(Y_t = j), \quad j = 1, \dots, K-1, \quad t = 1, \dots, T.$$

- L'approche consiste à se donner une modalité de référence, par exemple la modalité K, et à modéliser les prbabilités $p_j(x)$ selon

$$\log \frac{p_j(x_t)}{p_K(x_t)} = \beta_{1j} x_{t1} + \ldots + \beta_{pj} x_{tp} = x_t' \beta_j,$$

où $\beta_j = (\beta_{1j}, \dots, \beta_{pj})$.

Remarques

- Le modèle ne dépend pas de la modalité de référence choisie ! (seule la valeur des coefficients, et donc leur interprétation, en dépend).
- Ce modèle comprend $p(K-1)$ paramètres à estimer.
- Si $K = 2$, on retombe sur le modèle logistique binaire.

- On dispose d'une variable explicative Y à K modalités et on cherche à modéliser les probabilités

$$\mathbf{P}(Y_t = j), \quad j = 1, \dots, K-1, \quad t = 1, \dots, T.$$

- L'approche consiste à se donner une modalité de référence, par exemple la modalité K, et à modéliser les prbabilités $p_j(x)$ selon

$$\log \frac{p_j(x_t)}{p_K(x_t)} = \beta_{1j} x_{t1} + \ldots + \beta_{pj} x_{tp} = x_t' \beta_j,$$

où $\beta_j = (\beta_{1j}, \dots, \beta_{pj})$.

Remarques

- Le modèle ne dépend pas de la modalité de référence choisie ! (seule la valeur des coefficients, et donc leur interprétation, en dépend).
- Ce modèle comprend $p(K-1)$ paramètres à estimer.
- Si $K = 2$, on retombe sur le modèle logistique binaire.

Estimation des paramètres

- Tout se passe comme dans le cas binaire... mais les choses (vraisemblance, Information de Fisher...) sont beaucoup plus lourdes à écrire.
- Les paramètres inconnus du modèle sont estimés par maximum de vraisemblance et on montre que, sous des hypothèses similaires au cas binaire,

$$\sqrt{n}(\hat{\beta}-\beta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathcal{I}(\beta)^{-1}).$$

- On en déduit des procédures de tests (Wald, rapport de vraisemblance, score) ainsi que des intervalles de confiance pour les paramètres du modèle.

Estimation des paramètres

- Tout se passe comme dans le cas binaire... mais les choses (vraisemblance, Information de Fisher...) sont beaucoup plus lourdes à écrire.
- Les paramètres inconnus du modèle sont estimés par maximum de vraisemblance et on montre que, sous des hypothèses similaires au cas binaire,

$$\sqrt{n}(\hat{\beta}-\beta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathcal{I}(\beta)^{-1}).$$

- On en déduit des procédures de tests (Wald, rapport de vraisemblance, score) ainsi que des intervalles de confiance pour les paramètres du modèle.

Estimation des paramètres

- Tout se passe comme dans le cas binaire... mais les choses (vraisemblance, Information de Fisher...) sont beaucoup plus lourdes à écrire.
- Les paramètres inconnus du modèle sont estimés par maximum de vraisemblance et on montre que, sous des hypothèses similaires au cas binaire,

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathcal{I}(\beta)^{-1}).$$

- On en déduit des procédures de tests (Wald, rapport de vraisemblance, score) ainsi que des intervalles de confiance pour les paramètres du modèle.

Exemple

- Sur R, les fonctions **multinom** du package nnet ou **vglm** du package vgam permettent d'ajuster le modèle de régression logistique multinomial. Sous SAS, on pourra utiliser la **proc CATMOD**.

```
> model <- vglm(cbind(A,B,C)~business+punish+male+explain,data=donnees1,mu
> model1 <- multinom(wallet~business+punish+male+explain,data=donnees)
> model
Call:
vglm(formula = cbind(A, B, C) ~ business + punish + male + explain,
    family = multinomial, data = donnees1)

Coefficients:
(Intercept):1 (Intercept):2   business1:1   business1:2     punish2:1
   -2.4062097    -1.1068174     1.1791118     0.4155709     1.1450946
    punish2:2     punish3:1     punish3:2       male1:1       male1:2
    0.2491692     2.1411710     0.3531734     1.2672026     1.1716184
   explain1:1    explain1:2
   -1.5935358    -0.7978215

Degrees of Freedom: 46 Total; 34 Residual
Residual deviance: 31.91969
Log-likelihood: -46.32335
```

- On peut tester le modèle contre la constante ou l'effet de chaque variable à l'aide d'un test de rapport de vraisemblance :

```
> lrtest_vglm(model)
Likelihood ratio test

Model 1: cbind(A, B, C) ~ business + punish + male + explain
Model 2: cbind(A, B, C) ~ 1
  #Df  LogLik Df  Chisq Pr(>Chisq)
1  34 -46.323
2  44 -71.613 10 50.579  2.088e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> Anova(model1,type=3)
Analysis of Deviance Table (Type III tests)

Response: wallet
         LR Chisq Df Pr(>Chisq)
business   4.6540  2   0.097590 .
punish    11.0788  4   0.025692 *
male      13.0036  2   0.001501 **
explain    9.9181  2   0.007019 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Prévision

- Pour une nouvelle valeur de $x_{n+1} \in \mathbb{R}^p$, on peut naturellement estimer les probabilités que Y soit égales à j

$$p_j(x_{n+1}) = \mathbf{P}(Y_{n+1} = j) = \frac{\exp(x'_{n+1}\beta_j)}{1 + \sum_{j=1}^{K-1} \exp(x'_{n+1}\beta_j)}$$

par

$$\hat{p}_j(x_{n+1}) = \hat{\mathbf{P}}(Y_{n+1} = j) = \frac{\exp(x'_{n+1}\hat{\beta}_j)}{1 + \sum_{j=1}^{K-1} \exp(x'_{n+1}\hat{\beta}_j)}.$$

- Et connaissant la loi (asymptotique) des $\hat{\beta}_j$, on peut en déduire des intervalles de confiance pour $p_j(x_{n+1})$.

Remarque

Les notions de tests d'adéquation, de résidus, de critères de performance (AIC, BIC), de sélection de variables vues pour le cas binaire se généralisent "aisément" au modèle de régression logistique

Prévision

- Pour une nouvelle valeur de $x_{n+1} \in \mathbb{R}^p$, on peut naturellement estimer les probabilités que Y soit égales à j

$$p_j(x_{n+1}) = \mathbf{P}(Y_{n+1} = j) = \frac{\exp(x'_{n+1}\beta_j)}{1 + \sum_{j=1}^{K-1} \exp(x'_{n+1}\beta_j)}$$

par

$$\hat{p}_j(x_{n+1}) = \hat{\mathbf{P}}(Y_{n+1} = j) = \frac{\exp(x'_{n+1}\hat{\beta}_j)}{1 + \sum_{j=1}^{K-1} \exp(x'_{n+1}\hat{\beta}_j)}.$$

- Et connaissant la loi (asymptotique) des $\hat{\beta}_j$, on peut en déduire des intervalles de confiance pour $p_j(x_{n+1})$.

Remarque

Les notions de tests d'adéquation, de résidus, de critères de performance (AIC, BIC), de sélection de variables vues pour le cas binaire se généralisent "aisément" au modèle de régression logistique

Prévision

- Pour une nouvelle valeur de $x_{n+1} \in \mathbb{R}^p$, on peut naturellement estimer les probabilités que Y soit égales à j

$$p_j(x_{n+1}) = \mathbf{P}(Y_{n+1} = j) = \frac{\exp(x'_{n+1}\beta_j)}{1 + \sum_{j=1}^{K-1} \exp(x'_{n+1}\beta_j)}$$

par

$$\hat{p}_j(x_{n+1}) = \hat{\mathbf{P}}(Y_{n+1} = j) = \frac{\exp(x'_{n+1}\hat{\beta}_j)}{1 + \sum_{j=1}^{K-1} \exp(x'_{n+1}\hat{\beta}_j)}.$$

- Et connaissant la loi (asymptotique) des $\hat{\beta}_j$, on peut en déduire des intervalles de confiance pour $p_j(x_{n+1})$.

Remarque

Les notions de tests d'adéquation, de résidus, de critères de performance (AIC, BIC), de sélection de variables vues pour le cas binaire se généralisent "aisément" au modèle de régression logistique

Prévision

- Pour une nouvelle valeur de $x_{n+1} \in \mathbb{R}^p$, on peut naturellement estimer les probabilités que Y soit égales à j

$$p_j(x_{n+1}) = \mathbf{P}(Y_{n+1} = j) = \frac{\exp(x'_{n+1}\beta_j)}{1 + \sum_{j=1}^{K-1} \exp(x'_{n+1}\beta_j)}$$

par

$$\hat{p}_j(x_{n+1}) = \hat{\mathbf{P}}(Y_{n+1} = j) = \frac{\exp(x'_{n+1}\hat{\beta}_j)}{1 + \sum_{j=1}^{K-1} \exp(x'_{n+1}\hat{\beta}_j)}.$$

- Et connaissant la loi (asymptotique) des $\hat{\beta}_j$, on peut en déduire des intervalles de confiance pour $p_j(x_{n+1})$.

Remarque

Les notions de tests d'adéquation, de résidus, de critères de performance (AIC, BIC), de sélection de variables vues pour le cas binaire se généralisent "aisément" au modèle de régression logistique

Le modèle de régression logistique ordinal

Rappel sur le modèle logistique binaire

- On suppose pour simplifier qu'on dispose d'une seule variable explicative X.
- On suppose qu'il existe une variable latente (inobservée) $Y^\star$

$$Y_i^\star = \tilde{\beta}_0 + \beta_1 x_i + \varepsilon$$

où ε est une variable aléatoire centrée, telle que

$$Y_i = \mathbf{1}_{Y_i^\star > s}, \quad s \in \mathbb{R}.$$

- On a alors

$$\mathbf{P}(Y_i = 1) = \mathbf{P}(-\varepsilon < \beta_0 + \beta_1 x_i) = F_\varepsilon(\beta_0 + \beta_1 x_i)$$

où $\beta_0 = \tilde{\beta}_0 - s$.

Rappel sur le modèle logistique binaire

- On suppose pour simplifier qu'on dispose d'une seule variable explicative X.
- On suppose qu'il existe une variable latente (inobservée) $Y^\star$

$$Y_i^\star = \tilde{\beta}_0 + \beta_1 x_i + \varepsilon$$

où ε est une variable aléatoire centrée, telle que

$$Y_i = \mathbf{1}_{Y_i^\star > s}, \quad s \in \mathbb{R}.$$

- On a alors

$$\mathbf{P}(Y_i = 1) = \mathbf{P}(-\varepsilon < \beta_0 + \beta_1 x_i) = F_\varepsilon(\beta_0 + \beta_1 x_i)$$

où $\beta_0 = \tilde{\beta}_0 - s$.

Rappel sur le modèle logistique binaire

- On suppose pour simplifier qu'on dispose d'une seule variable explicative X.
- On suppose qu'il existe une variable latente (inobservée) $Y^\star$

$$Y_i^\star = \tilde{\beta}_0 + \beta_1 x_i + \varepsilon$$

où ε est une variable aléatoire centrée, telle que

$$Y_i = \mathbf{1}_{Y_i^\star > s}, \quad s \in \mathbb{R}.$$

- On a alors

$$\mathbf{P}(Y_i = 1) = \mathbf{P}(-\varepsilon < \beta_0 + \beta_1 x_i) = F_\varepsilon(\beta_0 + \beta_1 x_i)$$

où $\beta_0 = \tilde{\beta}_0 - s$.

- Définir le modèle revient à spécifier la loi de ε.

Propriété

- Si ε suit une loi logistique, c'est à dire de fonction de répartition

$$F_\varepsilon(x) = \frac{\exp(x)}{1+\exp(x)},$$

  alors le modèle est le modèle logistique.
- Si ε suit une loi normale centrée réduite alors le modèle est le modèle probit.

Remarque

On peut prolonger cette idée pour une variable Y ordinale à plus de deux modalités.

- Définir le modèle revient à spécifier la loi de ε.

Propriété

- Si ε suit une loi logistique, c'est à dire de fonction de répartition

$$F_\varepsilon(x) = \frac{\exp(x)}{1+\exp(x)},$$

alors le modèle est le modèle logistique.

- Si ε suit une loi normale centrée réduite alors le modèle est le modèle probit.

Remarque

On peut prolonger cette idée pour une variable Y ordinale à plus de deux modalités.

- Définir le modèle revient à spécifier la loi de ε.

Propriété

- Si ε suit une loi logistique, c'est à dire de fonction de répartition

$$F_\varepsilon(x) = \frac{\exp(x)}{1+\exp(x)},$$

alors le modèle est le modèle logistique.

- Si ε suit une loi normale centrée réduite alors le modèle est le modèle probit.

Remarque

On peut prolonger cette idée pour une variable Y ordinale à plus de deux modalités.

- Définir le modèle revient à spécifier la loi de ε.

Propriété

- Si ε suit une loi logistique, c'est à dire de fonction de répartition

$$F_\varepsilon(x) = \frac{\exp(x)}{1+\exp(x)},$$

  alors le modèle est le modèle logistique.
- Si ε suit une loi normale centrée réduite alors le modèle est le modèle probit.

Remarque

On peut prolonger cette idée pour une variable Y ordinale à plus de deux modalités.

- On cherche à expliquer Y ordinale à valeurs dans $\{1, \dots, K\}$ par p variables $X_1, \dots, X_p$. On dispose d'un n-échantillon indépendant $(x_1, Y_1), \dots, (x_n, Y_n)$.
- On suppose que Y_i est liée à une variable latente $Y_i^\star$ telle que

$$Y_i^\star = \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \varepsilon_i$$

selon

$$Y_i = \begin{cases} 1 & \text{si } Y_i^\star < \alpha_1 \\ j & \text{si } \alpha_{j-1} \leq Y_i^\star < \alpha_j, \ j = 2, \dots, K-1 \\ k & \text{si } Y_i^\star \geq \alpha_{K-1} \end{cases}$$

où $\alpha_1, \dots, \alpha_{K-1}$ sont des seuils inconnus et ε est un terme d'erreur aléatoire.

Le modèle logistique ordinal

Si les ε sont de loi logistique, on obtient alors le modèle logistique ordinal

$$\operatorname{logit} \mathbf{P}_\beta(Y_i \leq j) = \alpha_j - \beta_1 x_{i1} - \ldots - \beta_p x_{ip} = \alpha_j - x_i'\beta, \ j = 1, \dots, K-1.$$

- On cherche à expliquer Y ordinale à valeurs dans $\{1, \dots, K\}$ par p variables $X_1, \dots, X_p$. On dispose d'un n-échantillon indépendant $(x_1, Y_1), \dots, (x_n, Y_n)$.
- On suppose que Y_i est liée à une variable latente $Y_i^\star$ telle que

$$Y_i^\star = \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i$$

selon

$$Y_i = \begin{cases} 1 & \text{si } Y_i^\star < \alpha_1 \\ j & \text{si } \alpha_{j-1} \leq Y_i^\star < \alpha_j, \ j = 2, \dots, K-1 \\ k & \text{si } Y_i^\star \geq \alpha_{K-1} \end{cases}$$

où $\alpha_1, \dots, \alpha_{K-1}$ sont des seuils inconnus et ε est un terme d'erreur aléatoire.

Le modèle logistique ordinal

Si les ε sont de loi logistique, on obtient alors le modèle logistique ordinal

$$\operatorname{logit} \mathbf{P}_\beta(Y_i \leq j) = \alpha_j - \beta_1 x_{i1} - \dots - \beta_p x_{ip} = \alpha_j - x_i' \beta, \ j = 1, \dots, K-1.$$

Remarques

- Ce modèle est aussi appelé modèle cumulatif (car il prend en compte les logit des probabilités $\{Y \leq j\}$ ou encore modèle logistique à égalité des pentes (on verra plus tard pourquoi).
- Les coefficients β associés aux variables explicatives ne dépendent pas de j, seules les constantes α_j en dépendent.
- Le modèle nécessite l'estimation de $p + K - 1$ paramètres.
- La paramétrisation peut varier selon les logiciels, la `proc logistic` de `SAS` ajuste par exemple le modèle

$$\operatorname{logit} \mathbf{P}_\beta(Y_i \leq j) = \alpha_j + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} = \alpha_j + x_i'\beta,\ j = 1, \ldots, K - 1.$$

Remarques

- Ce modèle est aussi appelé modèle cumulatif (car il prend en compte les logit des probabilités $\{Y \leq j\}$ ou encore modèle logistique à égalité des pentes (on verra plus tard pourquoi).
- Les coefficients β associés aux variables explicatives ne dépendent pas de j, seules les constantes α_j en dépendent.
- Le modèle nécessite l'estimation de $p + K - 1$ paramètres.
- La paramétrisation peut varier selon les logiciels, la `proc logistic` de `SAS` ajuste par exemple le modèle

$$\mathrm{logit}\,\mathbf{P}_\beta(Y_i \leq j) = \alpha_j + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} = \alpha_j + x_i'\beta,\ j = 1, \ldots, K-1.$$

Remarques

- Ce modèle est aussi appelé modèle cumulatif (car il prend en compte les logit des probabilités $\{Y \leq j\}$ ou encore modèle logistique à égalité des pentes (on verra plus tard pourquoi).
- Les coefficients β associés aux variables explicatives ne dépendent pas de j, seules les constantes α_j en dépendent.
- Le modèle nécessite l'estimation de $p + K - 1$ paramètres.
- La paramétrisation peut varier selon les logiciels, la `proc logistic` de `SAS` ajuste par exemple le modèle

$$\text{logit}\, \mathbf{P}_\beta(Y_i \leq j) = \alpha_j + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} = \alpha_j + x_i'\beta, \; j = 1, \ldots, K-1.$$

Remarques

- Ce modèle est aussi appelé modèle cumulatif (car il prend en compte les logit des probabilités $\{Y \leq j\}$ ou encore modèle logistique à égalité des pentes (on verra plus tard pourquoi).
- Les coefficients β associés aux variables explicatives ne dépendent pas de j, seules les constantes α_j en dépendent.
- Le modèle nécessite l'estimation de $p + K - 1$ paramètres.
- La paramétrisation peut varier selon les logiciels, la `proc logistic` de `SAS` ajuste par exemple le modèle

$$\operatorname{logit} \mathbf{P}_\beta(Y_i \leq j) = \alpha_j + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} = \alpha_j + x_i'\beta, \; j = 1, \ldots, K-1.$$

Exemples

- Sous R, on peut utiliser la fonction **polr** du package MASS

```
> model <- polr(wallet~.,data=donnees)
> model
Coefficients:
     male1  business1    punish2    punish3   explain1
-1.0598227 -0.7388746 -0.6276423 -1.4030892  1.0518775
Intercepts:
       1|2        2|3
-2.5678520 -0.7890143
```

- ou la fonction **vglm** du package VGAM

```
> model1 <- vglm(cbind(A,B,C)~business+punish+male+explain,data=donne
                                                      cumulative(parallel=

> model1
Coefficients:
(Intercept):1  (Intercept):2      business1        punish2        punish3
   -2.5678316     -0.7889979      0.7388749      0.6276413      1.4030657
        male1       explain1
    1.0598180     -1.0518680
```

Moralité

Là encore, il convient d'aller voir dans l'aide la paramétrisation des fonctions.

Exemples

- Sous R, on peut utiliser la fonction **polr** du package MASS

```
> model <- polr(wallet~.,data=donnees)
> model
Coefficients:
     male1  business1    punish2    punish3   explain1
-1.0598227 -0.7388746 -0.6276423 -1.4030892  1.0518775
Intercepts:
       1|2        2|3
-2.5678520 -0.7890143
```

- ou la fonction **vglm** du package VGAM

```
> model1 <- vglm(cbind(A,B,C)~business+punish+male+explain,data=donne
                                                  cumulative(parallel=
> model1
Coefficients:
(Intercept):1 (Intercept):2     business1       punish2       punish3
   -2.5678316    -0.7889979     0.7388749     0.6276413     1.4030657
        male1      explain1
    1.0598180    -1.0518680
```

Moralité

Là encore, il convient d'aller voir dans l'aide la paramétrisation des fonctions.

Exemples

- Sous R, on peut utiliser la fonction **polr** du package MASS

```
> model <- polr(wallet~.,data=donnees)
> model
Coefficients:
     male1  business1    punish2    punish3   explain1
-1.0598227 -0.7388746 -0.6276423 -1.4030892  1.0518775
Intercepts:
       1|2        2|3
-2.5678520 -0.7890143
```

- ou la fonction **vglm** du package VGAM

```
> model1 <- vglm(cbind(A,B,C)~business+punish+male+explain,data=donne
                                                      cumulative(parallel=

> model1
Coefficients:
(Intercept):1 (Intercept):2     business1       punish2       punish3
   -2.5678316    -0.7889979     0.7388749     0.6276413     1.4030657
        male1      explain1
    1.0598180    -1.0518680
```

Moralité

Là encore, il convient d'aller voir dans l'aide la paramétrisation des fonctions.

Estimation des paramètres

- Les paramètres du modèle sont toujours estimés par maximum de vraisemblance et, sous des hypothèses similaires au cas binaire, on obtient la normalité asymptotique des estimateurs (on en déduit des IC et des procédures de test).

- Prévision : pour un nouvel individu x_{n+1} on pourra estimer les probabilités

$$\mathbf{P}_\beta(Y_{n+1} \leq j) = \frac{\exp(\alpha_j - x'_{n+1}\beta)}{1 + \exp(\alpha_j - x'_{n+1}\beta)}$$

par

$$\mathbf{P}_{\hat{\beta}}(Y_{n+1} \leq j) = \frac{\exp(\hat{\alpha}_j - x'_{n+1}\hat{\beta})}{1 + \exp(\hat{\alpha}_j - x'_{n+1}\hat{\beta})}$$

et en déduire ainsi les estimations des probabilités $\mathbf{P}_\beta(Y_{n+1} = j)$ pour faire la prévision.

Estimation des paramètres

- Les paramètres du modèle sont toujours estimés par maximum de vraisemblance et, sous des hypothèses similaires au cas binaire, on obtient la normalité asymptotique des estimateurs (on en déduit des IC et des procédures de test).
- Prévision : pour un nouvel individu x_{n+1} on pourra estimer les probabilités

$$\mathbf{P}_{\beta}(Y_{n+1} \leq j) = \frac{\exp(\alpha_j - x'_{n+1}\beta)}{1 + \exp(\alpha_j - x'_{n+1}\beta)}$$

par

$$\mathbf{P}_{\hat{\beta}}(Y_{n+1} \leq j) = \frac{\exp(\hat{\alpha}_j - x'_{n+1}\hat{\beta})}{1 + \exp(\hat{\alpha}_j - x'_{n+1}\hat{\beta})}$$

et en déduire ainsi les estimations des probabilités $\mathbf{P}_{\beta}(Y_{n+1} = j)$ pour faire la prévision.

Estimation des paramètres

- Les paramètres du modèle sont toujours estimés par maximum de vraisemblance et, sous des hypothèses similaires au cas binaire, on obtient la normalité asymptotique des estimateurs (on en déduit des IC et des procédures de test).
- Prévision : pour un nouvel individu x_{n+1} on pourra estimer les probabilités

$$\mathbf{P}_\beta(Y_{n+1} \leq j) = \frac{\exp(\alpha_j - x'_{n+1}\beta)}{1 + \exp(\alpha_j - x'_{n+1}\beta)}$$

par

$$\mathbf{P}_{\hat\beta}(Y_{n+1} \leq j) = \frac{\exp(\hat\alpha_j - x'_{n+1}\hat\beta)}{1 + \exp(\hat\alpha_j - x'_{n+1}\hat\beta)}$$

et en déduire ainsi les estimations des probabilités $\mathbf{P}_\beta(Y_{n+1} = j)$ pour faire la prévision.

Tests de rapport de vraisemblance

- Comme pour le modèle logistique multinomial, on peut tester le modèle contre la constante ou l'effet de chaque variable à l'aide de tests de rapport de vraisemblance :

```
> lrtest_vglm(model1)
Likelihood ratio test

Model 1: cbind(A, B, C) ~ business + punish + male + explain
Model 2: cbind(A, B, C) ~ 1
  #Df  LogLik Df  Chisq Pr(>Chisq)
1  39 -49.211
2  44 -71.613  5 44.805   1.59e-08 ***

> Anova(model,type=3)
Analysis of Deviance Table (Type III tests)

Response: wallet
         LR Chisq Df Pr(>Chisq)
male      10.9265  1   0.000948 ***
business   4.2667  1   0.038867 *
punish     9.1512  2   0.010300 *
explain    9.5168  1   0.002036 **
```

Proportionnalité des odd ratio

- Pour $x_i \in \mathbb{R}^p$, on définit

$$odd(x_i; Y \leq j \text{ vs } Y > j) = \frac{\mathbf{P}_\beta(Y_i \leq j)}{\mathbf{P}_\beta(Y_i > j)}.$$

- On a alors que l'odd ratio

$$OR(x_i, x_k; Y \leq j \text{ vs } Y > j) = \exp((x_k - x_i)'\beta),$$

ne dépend pas de j. Cette propriété est appélée proportionnalité des odd ratio.

- Les logiciels renvoient souvent les OR associés à ces évènements (qui sont rarement simples à interpréter ! Voir TP2).

Proportionnalité des odd ratio

- Pour $x_i \in \mathbb{R}^p$, on définit

$$odd(x_i;\, Y \leq j \text{ vs } Y > j) = \frac{\mathbf{P}_\beta(Y_i \leq j)}{\mathbf{P}_\beta(Y_i > j)}.$$

- On a alors que l'odd ratio

$$OR(x_i, x_k;\, Y \leq j \text{ vs } Y > j) = \exp((x_k - x_i)'\beta),$$

ne dépend pas de j. Cette propriété est appelée proportionnalité des odd ratio.

- Les logiciels renvoient souvent les OR associés à ces évènements (qui sont rarement simples à interpréter ! Voir TP2).

Proportionnalité des odd ratio

- Pour $x_i \in \mathbb{R}^p$, on définit

$$odd(x_i; Y \leq j \text{ vs } Y > j) = \frac{\mathbf{P}_\beta(Y_i \leq j)}{\mathbf{P}_\beta(Y_i > j)}.$$

- On a alors que l'odd ratio

$$OR(x_i, x_k; Y \leq j \text{ vs } Y > j) = \exp((x_k - x_i)'\beta),$$

  ne dépend pas de j. Cette propriété est appélée proportionnalité des odd ratio.

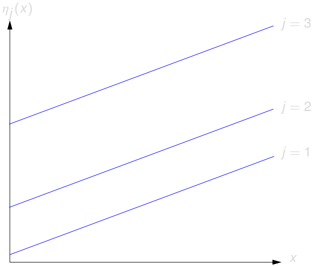
- Les logiciels renvoient souvent les OR associés à ces évènements (qui sont rarement simples à interpréter ! Voir TP2).

L'hypothèse d'égalité des pentes

- Supposons pour simplifier que l'on dispose d'une seule variable explicative X et considérons le modèle logistique ordinal

$$\operatorname{logit} \mathbf{P}_{\beta}(Y_i \leq j) = \alpha_j - \beta x_i.$$

- Seule la constante diffère selon j, c'est pourquoi on parle d'égalité des pentes.

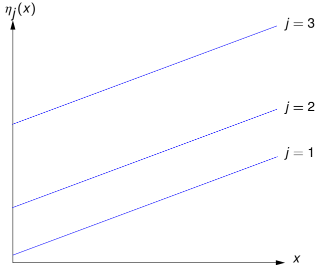


L'hypothèse d'égalité des pentes

- Supposons pour simplifier que l'on dispose d'une seule variable explicative X et considérons le modèle logistique ordinal

$$\operatorname{logit} \mathbf{P}_\beta(Y_i \leq j) = \alpha_j - \beta x_i.$$

- Seule la constante diffère selon j, c'est pourquoi on parle d'**égalité des pentes**.



- Si on lève cette propriété d'égalité des pentes (ce qui revient à considérer des pentes $\beta_j, j = 1, \ldots, K = 1$ différentes dans le modèle), on peut obtenir pour certaines valeurs de x_i

$$\mathbf{P}_\beta(Y_i \leq 1) > \mathbf{P}_\beta(Y_i \leq 2).$$

- Ce qui remet en cause le caractère ordinal de Y.
- Il est ainsi intéressant de développer un test permettant de vérifier l'égalité des pentes. On peut envisager des hypothèses du genre H_0 : "les pentes sont égales" contre H_1 : "elles ne le sont pas".

- Si on lève cette propriété d'égalité des pentes (ce qui revient à considérer des pentes $\beta_j, j = 1, \ldots, K = 1$ différentes dans le modèle), on peut obtenir pour certaines valeurs de x_i

$$\mathbf{P}_\beta(Y_i \leq 1) > \mathbf{P}_\beta(Y_i \leq 2).$$

- Ce qui remet en cause le caractère ordinal de Y.

- Il est ainsi intéressant de développer un test permettant de vérifier l'égalité des pentes. On peut envisager des hypothèses du genre H_0 : "les pentes sont égales" contre H_1 : "elles ne le sont pas".

- Si on lève cette propriété d'égalité des pentes (ce qui revient à considérer des pentes $\beta_j, j = 1, \dots, K = 1$ différentes dans le modèle), on peut obtenir pour certaines valeurs de x_i

$$\mathbf{P}_\beta(Y_i \leq 1) > \mathbf{P}_\beta(Y_i \leq 2).$$

- Ce qui remet en cause le caractère ordinal de Y.
- Il est ainsi intéressant de développer un test permettant de vérifier l'égalité des pentes. On peut envisager des hypothèses du genre H_0 : "les pentes sont égales" contre H_1 : "elles ne le sont pas".

Le test d'égalité des pentes

- Il suffit de voir que le modèle logistique ordinal (modèle à égalité de pentes) est emboité dans le modèle

$$\operatorname{logit} \mathbf{P}_\beta(Y_i \leq j) = \alpha_j - \beta_{1,j}x_{i1} - \ldots - \beta_{p,j}x_{ip}.$$

- Tester l'égalité des pentes revient donc à tester

$$H_0 : \beta_{\ell,1} = \ldots = \beta_{\ell,K-1}, \quad \forall \ell = 1, \ldots, p.$$

- On peut utiliser les statistiques de Wald, du rapport de vraisemblance ou du score pour effectuer ce test entre modèles emboités qui (pour n assez grand) suivent une loi χ^2 à

$$(K-1)(p-1) - (K-1+p) = p(K-2)$$

degrés de liberté.

Le test d'égalité des pentes

- Il suffit de voir que le modèle logistique ordinal (modèle à égalité de pentes) est emboité dans le modèle

$$\text{logit}\,\mathbf{P}_\beta(Y_i \leq j) = \alpha_j - \beta_{1,j}x_{i1} - \ldots - \beta_{p,j}x_{ip}.$$

- Tester l'égalité des pentes revient donc à tester

$$H_0 : \beta_{\ell,1} = \ldots = \beta_{\ell,K-1}, \quad \forall \ell = 1, \ldots, p.$$

- On peut utiliser les statistiques de Wald, du rapport de vraisemblance ou du score pour effectuer ce test entre modèles emboités qui (pour n assez grand) suivent une loi χ^2 à

$$(K-1)(p-1) - (K-1+p) = p(K-2)$$

degrés de liberté.

Le test d'égalité des pentes

- Il suffit de voir que le modèle logistique ordinal (modèle à égalité de pentes) est emboité dans le modèle

$$\mathrm{logit}\,\mathbf{P}_\beta(Y_i \leq j) = \alpha_j - \beta_{1,j}x_{i1} - \ldots - \beta_{p,j}x_{ip}.$$

- Tester l'égalité des pentes revient donc à tester

$$H_0 : \beta_{\ell,1} = \ldots = \beta_{\ell,K-1}, \quad \forall \ell = 1, \ldots, p.$$

- On peut utiliser les statistiques de Wald, du rapport de vraisemblance ou du score pour effectuer ce test entre modèles emboités qui (pour n assez grand) suivent une loi χ^2 à

$$(K-1)(p-1) - (K-1+p) = p(K-2)$$

degrés de liberté.

Exemple

- On teste l'égalité des pentes pour le jeu de données `portefeuille` :

```
> model2 <- vglm(cbind(A,B,C)~business+punish+male+explain,data=donnees1,
                                              cumulative(parallel=FALSE))
> statRV <- -2*(logLik(model1)-logLik(model2))
> 1-pchisq(statRV,df=length(coef(model2))-length(coef(model1)))
[1] 0.4053199
```

On accepte l'hypothèse d'égalité des pentes au seuil $\alpha = 5\%$.

Remarque

Sous SAS, la **proc logistic** utilise par défaut la statistique du score pour tester l'égalité des pentes (voir TP2).

Exemple

- On teste l'égalité des pentes pour le jeu de données `portefeuille` :

```
> model2 <- vglm(cbind(A,B,C)~business+punish+male+explain,data=donnees1,
                                        cumulative(parallel=FALSE))
> statRV <- -2*(logLik(model1)-logLik(model2))
> 1-pchisq(statRV,df=length(coef(model2))-length(coef(model1)))
[1] 0.4053199
```

On accepte l'hypothèse d'égalité des pentes au seuil $\alpha = 5\%$.

Remarque

Sous SAS, la **proc logistic** utilise par défaut la statistique du score pour tester l'égalité des pentes (voir TP2).

Exemple

- On teste l'égalité des pentes pour le jeu de données `portefeuille` :

```
> model2 <- vglm(cbind(A,B,C)~business+punish+male+explain,data=donnees1,
                                         cumulative(parallel=FALSE))
> statRV <- -2*(logLik(model1)-logLik(model2))
> 1-pchisq(statRV,df=length(coef(model2))-length(coef(model1)))
[1] 0.4053199
```

On accepte l'hypothèse d'égalité des pentes au seuil $\alpha = 5\%$.

Remarque

Sous `SAS`, la **proc logistic** utilise par défaut la statistique du score pour tester l'égalité des pentes (voir TP2).

Remarque

Les notions de tests d'adéquation, de résidus, de critères de performance (AIC, BIC), de sélection de variables vues pour le cas binaire se généralisent "aisément" au modèle de régression logistique multinomial.

Exemple

En présence de données répétées, sous le modèle logistique ordinal, la déviance suit (pour n_t assez grand), une loi du χ_2 à $T(K-1)-(p+K-1)$ ddl.

- Exemple : Tests d'adéquation de déviance et de Pearson pour le `model1`.

```
> deviance(model1) #statistique de la déviance
[1] 37.69427
> 1-pchisq(deviance(model1),df=39)
[1] 0.5293916
> model1@res.ss #statistique de Pearson
[1] 31.41399
> 1-pchisq(model1@res.ss,df=39)
[1] 0.8009696
```

Remarque

Les notions de tests d'adéquation, de résidus, de critères de performance (AIC, BIC), de sélection de variables vues pour le cas binaire se généralisent "aisément" au modèle de régression logistique multinomial.

Exemple

En présence de données répétées, sous le modèle logistique ordinal, la déviance suit (pour n_t assez grand), une loi du χ_2 à $T(K-1)-(p+K-1)$ ddl.

- Exemple : Tests d'adéquation de déviance et de Pearson pour le `model1`.

```
> deviance(model1) #statistique de la déviance
[1] 37.69427
> 1-pchisq(deviance(model1),df=39)
[1] 0.5293916
> model1@res.ss #statistique de Pearson
[1] 31.41399
> 1-pchisq(model1@res.ss,df=39)
[1] 0.8009696
```

Remarque

Les notions de tests d'adéquation, de résidus, de critères de performance (AIC, BIC), de sélection de variables vues pour le cas binaire se généralisent "aisément" au modèle de régression logistique multinomial.

Exemple

En présence de données répétées, sous le modèle logistique ordinal, la déviance suit (pour n_t assez grand), une loi du χ_2 à $T(K-1)-(p+K-1)$ ddl.

- Exemple : Tests d'adéquation de déviance et de Pearson pour le `model1`.

```
> deviance(model1) #statistique de la déviance
[1] 37.69427
> 1-pchisq(deviance(model1),df=39)
[1] 0.5293916
> model1@res.ss #statistique de Pearson
[1] 31.41399
> 1-pchisq(model1@res.ss,df=39)
[1] 0.8009696
```

Références I