Die partielle Likelihood-Funktion

Roger Züst

12. Juni 2006

1 Repetition: Maximum-Likelihood-Methode

Hat man n unabhängige Beobachtungen $x_1, x_2, \ldots, x_n$ einer Zufallsvariablen X und eine Familie von möglichen Dichten f_θ $(\theta = (\theta_1, \ldots, \theta_q) \in \Theta \subset \mathbb{R}^q)$, dann ist die Likelihood Funktion gegeben durch

$$L(\theta) = \prod_{i=1}^{n} f_\theta(x_i).$$

Der Maximum-Likelihood-Schätzer ist definiert als

$$\hat{\theta}(x) = arg \max_\theta L(\theta) = arg \max_\theta log(L(\theta)).$$

Unter geeigneten Regularitätsbedingungen ist $\hat{\theta}(x)$ eindeutig definiert durch die Gleichungen

$$0 = \frac{\partial}{\partial \theta_k} log(L(\theta)) = \sum_{i=1}^{n} \frac{\partial}{\partial \theta_k} log(f_\theta(x_i)), \quad \forall k = 1, \ldots, q.$$

2 Einführung

Im Folgenden ist die konditionierte Risikofunktion gegeben durch

$$\lambda(t|Z) = \lambda_0(t) exp(\beta' Z),$$

wobei, falls nicht anders erwähnt, die Kovarianten Z als zeitunabhängig angenommen werden. Es gilt die unbekannten Parameter λ_0 und β zu schätzen. Da aber λ_0 ein unendlich dimensionaler Parameter ist lässt sich die Standard Maximum-Likelihood-Methode nicht anwenden. Das Ziel ist, falls möglich, die Regressionsparameter β von der Basisfunktion λ_0 loszulösen, um sie unabhängig von λ_0 schätzen zu können.

3 Notation

Ist nichts anderes erwähnt, sind die beobachteten Zeiten zu verschieden Ereignissen verschieden. Es sind die unabhängigen Tripel $\{(X_i, \delta_i, Z_i) : i = 1, \ldots, n\}$ und die beobachteten Ausfallzeiten $T_1^O < T_2^O < \ldots < T_L^O$ gegeben (man setzt $T_0^O = 0$ und $T_{L+1}^O = \infty$). Dabei ist (i) der Proband der zum Zeitpunkt T_i^O ausfällt. Es bezeichne m_i die Anzahl Probanden die im Intervall $[T_i^O, T_{i+1}^O)$ zensiert werden, dazu sind (i,j) $(j = 1, \ldots, m_i)$ die enstprechenden Probanden.
Definiere $R_i = \{(j), (j,k) : j = i, \ldots, L, k = 1, \ldots, m_j\}$ die Risikomenge zum Zeitpunkt T_i^O.

Wir wollen zeigen, dass durch

$$L(\beta) = \prod_{i=1}^{L} \frac{exp(\beta' Z_{(i)})}{\sum_{l \in R_i} exp(\beta' Z_l)}$$

eine sinnvolle Likelihood-Funktion für β gegeben ist. Was ist speziell an dieser Funktion? Hier die wichtigsten Punkte:

1. Sie ist unabhängig von der Basisfunktion λ_0. Deshalb nennt man sie die **partielle Likelihood-Funktion (PLF)** für β.

2. Sie hängt nicht ab von den Zeitintervallen zwischen den Ereignissen, sondern nur von deren Reihenfolge.

3. Zensierte Probanden tragen nur über die Risikomenge zur Funktion bei.

4 PLF ohne Zensuren

Die Daten $\{(X_i, \delta_i, Z_i) : i = 1, \ldots, n\}$ sind unzensiert, das heisst $X_i \equiv T_i$, beziehungsweise $\delta_i \equiv 1$. Bezeichne mit $O = (T_1^0, T_2^0, \ldots, T_n^0)'$ die Ordnungsstatistik und mit $r = ((1), (2), \ldots, (n))'$ die Rangstatistik zu den beobachteten Zeiten. Zum Beispiel: Falls $(T_1, T_2, T_3, T_4) = (28, 15, 17, 6)$, dann ist $O = (6, 15, 17, 28)'$ und $r = (4, 2, 3, 1)'$.

G bezeichne die Gruppe der monoton steigenden Diffeomorphismen von $(0, \infty)$ auf sich selbst. Setzt man $V \equiv g^{-1}(T)$ für ein $g \in G$, dann sieht man durch eine Variablentransformation, dass die Risikofunktion von V gegeben ist durch

$$\begin{aligned} \lambda(v|Z) &= \lambda_0(g(v)) \frac{dg(v)}{dv} exp(\beta' Z) \\ &= \lambda_1(v) exp(\beta' Z), \end{aligned}$$

wobei λ_1 eine unbekannte Funktion ist. Somit ist unter $\{(T_i, Z_i) : i = 1, \ldots, n\}$ und $\{(V_i, Z_i) : i = 1, \ldots, n\}$ das Schätzproblem für β das gleiche. Mit anderen Worten: Der Schätzer für β sollte invariant sein unter G. Man kann mittels G eine Realisierung von O in jede andere überführen, wobei r unverändert bleibt. Das heisst einzig die Rangstatistik r ist entscheidend für die Schätzung von β.

Proposition 4.1. *Es gilt*

$$P\{r = ((1), \ldots, (n)); \beta\} = P\{T_{(1)} < \ldots < T_{(n)}; \beta\} = \prod_{i=1}^{n} \frac{exp(\beta' Z_{(i)})}{\sum_{l \in R_i} exp(\beta' Z_l)}$$

(wobei $R_i = \{(i), (i+1), \ldots, (n)\}$*).*

In diesem Fall ist $L(\beta)$ identisch mit der Likelihood-Funktion bezüglich der Rangstatistik.

5 PLF mit Zensuren

Der Unterschied zum vorherigen Kapitel ist, dass $\delta_i \equiv 1$ nicht mehr gelten muss. Ähnlich wie oben bezeichne $O = (X^O_{[1]}, X^O_{[2]}, \ldots, X^O_{[n]})'$ und $r = ([1], [2], \ldots, [n])'$ die Ordnungs-bzw. Rangstatistik zu den beobachteten Zeiten. Zusätzlich hat man aber noch die Information, ob es sich bei den Ereignissen um Ausfälle oder Zensuren handelt, also $\delta = (\delta_{[1]}, \delta_{[2]}, \ldots, \delta_{[n]})'$.

Wie oben bleiben r und δ unter Transformationen aus G unverändert, während sie transitiv auf O operieren. Allerdings ist die analoge Wahrscheinlichkeit wie die in Proposition 4.1, also

$$P\{X_{[1]} < \ldots < X_{[n]}; \beta\} = P\{min(T_{[1]}, U_{[1]}) < \ldots < min(T_{[n]}, U_{[n]}); \beta\}$$

schwierig zu berechnen, da man nun neben der Verteilung von T auch die von U berücksichtigen muss.

Man kann folgenden Ausweg wählen: Man berechnet nicht mehr die Wahrscheinlichkeit einer spezifischen Rangfolge r, sondern die aller Rangfolgen, die konsistent zu r sind. Dabei sind genau diese Rangfolgen konsistent zu r, die man hätte beobachten können wären die zensierten Probanden zu irgend einem späteren Zeitpunkt ausgefallen. Zur Veranschaulichung sei $r = (4, 2, 3, 1)'$ und $\delta = (1, 0, 1, 0)'$, dann sind $(4, 2, 3, 1)'$, $(4, 3, 2, 1)'$ und $(4, 3, 1, 2)'$ alle konsistenten Rangfolgen zu r.

Proposition 5.1. *Es gilt*

$$P\{r \text{ ist konsistent zu } ([1], [2], \ldots, [n]); \beta\} =$$

$$P\{T_{(1)} < \ldots < T_{(L)}, T_{(i)} < T_{(i,j)}; i = 1, \ldots, L; j = 1, \ldots, m_i; \beta\} = \prod_{i=1}^{L} \frac{exp(\beta' Z_{(i)})}{\sum_{l \in R_i} exp(\beta' Z_l)}$$

(wobei $R_i = \{(j), (j,k) : j = i, \ldots, L, k = 1, \ldots, m_j\}$*).*

6 PLF mit gleichzeitigen Ereignissen

Wie bereits erwähnt ist eine Eigenschaft der partiellen Likelihood-Funktion, dass die Länge der Zeitintervalle zwischen den Ereignissen keine Rolle spielt. Einzig die Reihenfolge ist entscheidend. Durch Messfehler kann man sich möglicherweise der Reihenfolge nicht sicher sein und man möchte gewisse Ereignisse als gleichzeitig behandeln.

Wie oben sind $T_1^O < T_2^O < \ldots < T_L^O$ die beobachteten Ausfallzeiten. Zusätzlich zu den bekannten Bezeichnungen sei d_i die Anzahl Ausfälle zum Zeitpunkt T_i^O und $(i)^j, j = 1, \ldots, d_i$ die entsprechenden Probanden.

Mit ähnlichen Berechnungen wie zu den beiden Propositionen oben oder mit etwas allgemeineren Betrachtungen zur Konstruktion der partiellen Likelihood-Funktion kann man zeigen, dass falls d_i klein ist im Vergleich zur Risikomenge R_i, die Funktion

$$L(\beta) = \prod_{i=1}^{L} \frac{exp(\beta' \sum_{j=1}^{d_i} Z_{(i)^j})}{[\sum_{l \in R_i} exp(\beta' Z_l)]^{d_i}}$$

$$(R_i = \{(j)^l, (j,k) : j = i, \ldots, L, l = 1, \ldots, d_j, k = 1, \ldots, m_j\})$$

eine vernünftige Verallgemeinerung der partiellen Likelihood-Funktion aus Proposition 5.1 ist, selbst wenn man die Kovarianten Z_i als zeitabhängig annimmt.

7 Tests

$L_x(\beta)$ ist eine (partielle) Likelihood-Funktion bezüglich des Parameters $\beta = (\beta_1, \ldots, \beta_p)' \in \Theta$ und $\hat{\beta}$ der durch die Maximum-Likelihood-Methode geschätze Parameter. Der Score-Vektor $U(\beta)$ und die Informationsmatrix $I(\beta)$ werden definiert durch

$$U_i(\beta) = \frac{\partial}{\partial \beta_i} log(L(\beta)), \quad i = 1, \ldots, p$$

$$I_{ij}(\beta) = -\frac{\partial^2}{\partial \beta_i \partial \beta_j} log(L(\beta)), \quad i, j = 1, \ldots, p.$$

Ein üblicher Test für die Hypothese $\beta \in \Theta_0 \subset \Theta$ ist der **Likelihood-Quotienten-Test**, der die Hypothese verwirft, wenn die Teststatistik

$$T_{LQ}(x) = \frac{sup_{\beta \in \Theta} L_x(\beta)}{sup_{\beta \in \Theta_0} L_x(\beta)}$$

zu gross ist. In "regulären" Fällen ist $2logT_{LQ}(x)$ genähert χ_d^2-verteilt, wobei d die Dimensionsdifferenz zwischen Θ und Θ_0 bezeichnet.

Beispiel 7.1. Man ist in einer Testreihe an der Wirksamkeit eines Medikaments ($Z = 1$) gegenüber einem Placebo ($Z = 0$) interessiert. Die Nullhypothese lautet in diesem Fall $\beta = 0$ (d.h. das Medikament hat keine Wirkung). $2logT_{LQ}(x)$ ist für grosse n näherungsweise χ_1^2-verteilt.

Es seien noch zwei weitere Tests für die Nullhypothese $H_0 : \beta = \beta_0$ erwähnt. Der **Wald Test** und der **Rao Score Test** sind durch die Teststatistiken $T_W = (\hat{\beta}-\beta_0)' I(\hat{\beta})(\hat{\beta}-\beta_0)$ respektive $T_{RS} = U(\beta_0)' I^{-1}(\beta_0) U(\beta_0)$ gegeben. Beide sind für grosse n näherungsweise χ_p^2-verteilt.

8 Schätzung von $\Lambda(t|Z)$

Wie wir wissen, ist der Nelson Schätzer der kumulativen Hazardfunktion gegeben durch

$$\hat{\Lambda}(t) = \int_0^t \frac{\sum_{i=1}^n dN_i(s)}{\sum_{i=1}^n Y_i(s)}.$$

Wir wollen nun mit dessen Hilfe einen Schätzer für $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$ finden. Dazu sei $\hat{\beta}$ eine Schätzung der Regressionskoeffizienten. Der i-te Proband hat somit eine geschätzte Hazardfunktion von $exp(\hat{\beta}' Z_i)\lambda_0(s)$.

Im Nelson Schätzer haben wir $\sum_{i=1}^n Y_i(s)$ aktive Probanden, die alle die gleiche Ausfallrate $\lambda(s)$ besitzen. Nun ist jedoch $\lambda(s|Z)$ für verschiedene Probanden verschieden. Es gilt aber, dass ein Proband mit Ausfallrate $exp(\hat{\beta}' Z)\lambda_0(s)$ mit der gleichen Wahrscheinlichkeit ausfällt wie "$exp(\hat{\beta}' Z)$ Probanden", welche die selbe Ausfallrate $\lambda_0(s)$ besitzen. Man kann somit sagen, dass wir $\sum_{i=1}^n Y_i(s) exp(\hat{\beta}' Z_i)$ aktive Probanden mit Ausfallrate $\lambda_0(s)$ haben. Die entsprechende Anwendung des Nelson Schätzer auf Λ_0 ist somit durch

$$\hat{\Lambda}_0(t) = \int_0^t \frac{\sum_{i=1}^n dN_i(s)}{\sum_{i=1}^n Y_i(s) exp(\hat{\beta}' Z_i)}$$

gegeben.

Im Falle von zeitunabhängigen Kovarianten gilt

$$\Lambda(t|Z) = \int_0^t \lambda_0(t) exp(\beta' Z) ds = \Lambda_0(s) exp(\beta' Z).$$

Somit ist

$$\hat{\Lambda}(t|Z) = \hat{\Lambda}_0(s) exp(\hat{\beta}' Z)$$

ein Schätzer für die kumulative Hazardfunktion.