Proportionaler Hazard und Multiplikatives Dichtemodell

Nicolas Städler

June 11, 2006

Immer spielen in medizinischen Studien neben Ausfallszeiten und Zensurierungszeiten auch andere Zufallsvariabeln eine Rolle. Zum Beispiel stellt man sich in einer Studie die Frage ob die Verabreichung eines Medikaments einen Einfluss auf den Verlauf einer Krankheit hat. Man vergleicht dabei Patienten mit Placebo (Z=0) und Patienten mit Medikament (Z=1). Solche Zufallsvariablen, die einen Einfluss auf die Verteilung des Ereignisses haben, nennt man *Kovarianten*. Kovarianten lassen sich grob in 3 Gruppen gliedern: demographische Variablen (Alter, Geschlecht, Bildung,usw), verhaltens Variablen (Alkoholkonsum, Physischer Level,usw) und physiologische Variablen (Blutdruck, Hämoglobinwerte,usw). Zudem sind sie quantitativer (z.B. Alter) oder qualitativer Natur (z.B. Placebo versus Medikament).

In unserem Vortrag geht es als erstes darum Modelle zu entwickeln die den Einfluss einer Kovarianten auf die Ausfallszeit beschreiben. Zweitens werden wir eine Art von Likelihood für solche Modelle suchen und drittens werden wir Schätzer für solche Likelihoods vorstellen.

Es gibt zwei verschiedene Ansätze ein Modell mit einem p-dim Kovariantenvektor $Z = (Z_1, ..., Z_p)$ zu modellieren. Beide Ansätze werden semiparametrisch sein, d.h. das Modell wird nicht vollständig parametrisch sein. Es wird immer einen ∞-dimensionalen Teil geben, den sog. Baseline Hazard.

1 Proportionaler Hazard

Ausgehend von

$$N(t) = I_{[X \leq t, \delta = 1]} \qquad Y(t) = I_{[X \geq t]}$$

modelliert man bei der ersten Variante die *bedingte Hazard* Funktion:

$$\lambda(t|Z) = \lim_{h \to 0+} h^{-1} P(t \leq T < t + h | T \geq t, Z)$$

Wir nehmen zusätzlich an, dass für zwei verschiedene Kovariantenwerte Z_1 und Z_2, die dazugehörigen bedingten Fehlerraten ein zeitlich konstantes Verhältnis haben.

$$\lambda(t|Z_1) = k(Z_1, Z_2)\lambda(t|Z_2)$$

bezeichnen wir die Fehlerrate gegeben Z = 0 mit $\lambda_0(t)$ erhalten wir

$$\lambda(t|Z) = \lambda_0(t)g(Z)$$

wobei $g(Z) = k(Z, 0)$

Da wir eine möglichst einfache Abhängigkeit der Funktion g von Z wollen und $\lambda(t|Z)$ grösser als Null sein soll, erhalten wir unser *'Proportional hazard model'*:

$$\lambda(t|Z) = \lambda_0(t)exp(\beta Z)$$

β ist ein p-dim Regressionsvektor, ihn wollen wir später schätzen. λ_0 ist der *'Basline Hazard'*. Er ist derjenige Teil im Modell, der nicht parametrisch ist.

2 Multiplikatives Dichte Modell

Eine andere, elegante Methode der Modellierung beruht auf der Doob-Meyer Zerlegung. Mit ihr können allgemeinere Zählprozesse und zeitabhängige Kovarianten betrachtet werden. Beispielsweise will man auch wiederkehrende Ereignisse analysieren, wie Infektionen die mehrmals bei einem Patienten auftreten können.

Ausgehend von einem Zählprozess N, einem Zensurierungsprozess Y (1 falls Obijekt unter Beobachtung, 0 sonst) ,einem Kovariantenprozess Z und einer Geschichte $F_t = \sigma(Z(u+), N(u), Y(u+) : 0 \leq u \leq t)$ betrachten wir die Doob-Meyer Zerlegung:

$M = N - A$ A ist der previsible Prozess oder auch kumulative Dichteprozess

Heuristisch erhalten wir:

$$E(dN(s)|F_{s-}) = dA(s) \text{ und } P(dN(s) = 1|F_{s-}) = dA(s)$$

$dA(s)$ beschreibt nun in welcher Weise $dN(s)$ von der bisherigen Information über das Experiment abhängt. Wir möchten nun diese bedingte Sprungrate schreiben als $dA(s) = l(s)ds$. Die Idee ist nun dieses $dA(s) = l(s)ds$, bzw $l(s)$ zu modellieren. Man nennt den Prozess $l(s)$ auch *Dichteprozess* für $N(t)$. Ein einfaches Modell wäre nun $l(s) = \lambda_0(s)exp(Z(s))Y(s)$ zu setzen. Dieses stimmt für konstante Kovarianten und unabhängige Zensurierung mit dem Proportionalen Hazard überein.

Definition 1 *(Multiplikative Dichte Modell) Das Multiplikative Dichte Modell für Beobachtungen aus einer Population mit n Obijekten besteht aus einem Tripel,* (N_i, Y_i, Z_i)*,* $i = 1, ..., n$*, bestehend aus Zähl-, Zensurierungs- und Kovariantenprozess, einer rechtsstetigen Geschichte* F_t *und n Dichteprozessen* $l_i = \lambda_0 Y_i exp(\beta Z_i),$*,* $i = 1, ..., n$*. Das Modell muss zudem folgenden Bedingungen genügen:*

1. $N = (N_1, ..., N_n)$ *ist ein multivarianter Zählprozess*

2. $N_i - A_i$ *ist ein lokales Martingal bezüglich* F_t*, wobei* A_i *der stetige Kompensator* $A_i = \int l_i(s)\, ds$ *ist*

3. Y_i *und* Z_i *sind previsibel bezüglich* F_t *und die Prozesse* Z_i *sind lokal beschränkt*

Bemerkungen:

- $N = (N_1, ..., N_n)$ sorgt mit der Stetigkeit von A_i dass $d\langle M_i, M_j\rangle = 0$
- Diese allgemeine Definition sollte trotzdem eine klare Interpretation bezüglich unseren Daten liefern. Diese wurde mit den oben genannten Heuristiken bereits angedeutet. Folgendes Theorem bestätigt sie:

Theorem 1 *Sei N ein Zählprozess, A Kompensator bezüglich der Geschichte* F_t*. Nehme an* $A(t) = \int l_i\, ds$*. l(t) sei zudem linksstetig mit rechtsseitigen Limiten und beschränkt durch eine integrable Zufallsvariable Q. Dann gilt:*

1.

$$\lim_{h \to 0+} h^{-1} E(N(t+h) - N(t)|F_t) = l(t+) \qquad \textit{und}$$

2.

$$\lim_{h \to 0+} h^{-1} P(N(t+h) - N(t) = 1|F_t) = l(t+)$$

Ein Modellierer der ein Multiplikatives Dichte Modell auf einen Datensatz anwenden will muss sich einerseits vergewissern ob es wirklich eine Doob-Meyer Zerlegung bezüglich der Geschichte gibt, andererseits muss er begründen wieso der Kompensator von der Form $A = \int l$ ist. Die Theoreme 4.2.2 und 4.4.3 in 'Fleming und Harrington' beantworten die beiden Fragen unter zusätzlichen Annahmen mit ja.

Leider bleibt immer noch unklar, ob der Dichteprozess $l(s)$ (= bedingte Sprungrate) wirklich auch der statistisch interessante Parameter ist. Als Illustration ein **Beispiel**:

Seien T und U unabhängig. Der Effekt einer zeitunabhängigen Kovariante Z auf den Hazard sei gegeben durch: $\lambda(t|Z) = \lambda_0(t) exp(\beta Z)$.

Nach dem Theorem 1.3.1 folgt, dass

$$dA(t) = Y(t)\lambda_0(t) exp(\beta Z) dt.$$

Ein Modellierer gehe fälschlicherweise von einer homogenen Population aus. Sein Multiplikatives Dichtemodell basiert auf der Geschichte $\widetilde{F_t} = \sigma(N(u), N^U(u) : 0 \le u \le t)$. Damit erhält man folgende Dichtefunktion:

$$d\widetilde{A}(t) = \lambda_0(t) Y(t) E(exp(\beta Z|Y(t) = 1) dt$$

Sein *beobachteter* Hazard $\lambda_0(t) E(exp(\beta Z|Y(t) = 1)$ ist abhängig von β und von der bedingten Verteilung von Z, gegeben $Y(t) = 1$.