

# Tests im nichtparametrischen additiven Hazards Model

Boris Wälchli

06-26-2006

## 1 Einleitung

Die Schätzer der Regressionsfunktionen  $\beta_k(t)$  sollen auf ihren Einfluss auf die Hazard Rate untersucht werden. Konkret soll die Hypothese

$$H_0 : \beta_k(t) = 0 \quad \forall t \text{ und } \forall k \in K \text{ gegen } H_A : \exists \beta_k(t) \neq 0$$

getestet werden. Wobei  $K \subset \{1, \dots, p\}$  ist, also eine Teilmenge der Schätzer oder alle Schätzer beschreibt.

Wie beim Log Rank Test soll die geschätzte Rate mit der erwarteten Rate verglichen werden, wobei die erwartete Rate hier null ist. Gewichte sollen ebenfalls benutzt werden.

## 2 Teststatistik

Damit die Teststatistik konstruiert werden kann, muss zuerst die Gewichtsfunktion bestimmt werden. Sei  $H_{0k} : \beta_k(t) = 0$  eine von  $p$  möglichen Unterhypothesen und  $W_k(t)$  die Gewichtsfunktion für den Test der  $k$ -ten Unterhypothese.

Aalen <sup>1</sup> schlägt in seiner Arbeit zwei Gewichtsfunktionen vor.

---

<sup>1</sup>Aalen OO 1993. *Further Results on the nonparametric linear regression model in survival analysis*. *Statistics in Medicine* 1993; 12:1569-1588

Der erste Vorschlag beschreibt eine Gewicht, das zeitabhängig ist, aber für alle Regressionsfunktionen gleich:

$$W_k(t) = \sum_{j=0}^n Y_j(t)$$

Diese Funktion entspricht der Anzahl der beobachteten (gefährdeten) Personen zur Zeit  $t$ .

Der zweite Vorschlag nimmt für die unterschiedlichen Unterhypothesen verschiedene Werte an:

$$W_k(t) = \frac{1}{R_{k+1,k+1}}$$

Dabei ist  $R_{k+1,k+1}$  das  $k+1$  Diagonalelement der inversen Matrix der Matrix  $A$  aus der vorhergehenden Sektion.

Bhattacharyya und Klein <sup>2</sup> zeigen jedoch, dass die Wahl einer Gewichtsfunktion mit unterschiedlichen Werten für die verschiedene Unterhypothesen zu inkonsistenten Resultaten führt, sollte man Tests mit mehr als zwei Gruppen durchführen.

Die Teststatistik für  $H_{0k}$ , mit einer allgemeinen Gewichtsfunktion  $W_k$  und den Notationen aus der vorhergehenden Sektion, ist:

$$U_k = \sum_{j=0}^n \delta_j W_k(T_j) \sum_{h=1}^{p+1} R_{k+1,h}(T_j) Z_{j,k-1}(T_j) \quad k = 0, \dots, p$$

Wobei unter  $H_0$  asymptotisch gilt:

$$\frac{U_k}{\sqrt{\sigma_{kk}^2}} \stackrel{as}{\approx} N(0, 1)$$

$H_0$  wird verworfen, sollte  $\frac{U_k}{\sqrt{\sigma_{kk}^2}}$  signifikant von 0 abweichen.

Für diese Teststatistik ist eine Schätzung der Varianz von  $U_k$  nötig.

$$\sigma_{kk}^2 = \sum_{j=0}^n \delta_j W_k(T_j)^2 \sum_{h=1}^{p+1} \sum_{m=1}^{p+1} R_{k+1,m}(T_j) R_{k+1,n}(T_j) Z_{j,m-1}(T_j) Z_{j,n-1}(T_j) \quad k = 0, \dots, p$$

---

<sup>2</sup>Bhattacharyya, Mouchumi/ Klein, John P. (2005) *A note on testing in Aalen's additive hazards regression models* Statistics in Medicine 2005; 24:2235-2240

## 2.1 Mehrdimensionale Tests

Sollen  $q$  Unterhypothesen  $H_{0k} : \beta_k(t) = 0 \forall k = 1, \dots, q \leq p$  simultan getestet werden, muss die Kovarianzmatrix der Teststatistiken erzeugt werden. Die geschätzte Kovarianz von  $U_k$  und  $U_g$  ist

$$\sigma_{gk}^2 = \sum_{j=0}^n \delta_j W_g(T_j) W_k(T_j)^2 \sum_{h=1}^{p+1} \sum_{m=1}^{p+1} R_{g+1,m}(T_j) R_{k+1,n}(T_j) Z_{j,m-1}(T_j) Z_{j,n-1}(T_j)$$

$$g, k = 0, \dots, q$$

Somit sind die Kovarianzmatrix  $\mathbf{V}$  und der  $q$ -Vektor  $\mathbf{U}_q$  der  $q$  Teststatistiken der Unterhypothesen

$$\mathbf{V}_q = [\sigma_{gk}]_{g=1, \dots, q, k=1, \dots, q}$$

$$\mathbf{U}_q = \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_q \end{pmatrix}$$

Die Teststatistik ist dann die Quadratische Form

$$\mathbf{U}_q^t \mathbf{V}_q^{-1} \mathbf{U}_q,$$

wobei unter  $H_0$  asymptotisch gilt:

$$\mathbf{U}_q^t \mathbf{V}_q^{-1} \mathbf{U}_q \stackrel{as}{\sim} \chi_q^2$$

## 2.2 Anwendung

Das Zwei-Stichprobenbeispiel soll fortgesetzt werden. Für die Gewichtfunktion soll der zweite Vorschlag von Aalen verwendet werden.  $R$  ist wie zuvor :

$$R = \begin{pmatrix} \frac{1}{Y_2(t)} & \frac{-1}{Y_2(t)} \\ \frac{-1}{Y_2(t)} & \frac{1}{Y_1(t)} + \frac{1}{Y_2(t)} \end{pmatrix}$$

$$\Rightarrow W_1(t) = \frac{1}{R_{22}(t)} = \left( \frac{1}{Y_1(t)} + \frac{1}{Y_2(t)} \right)^{-1} = \frac{Y_1(t)Y_2(t)}{Y_1(t)+Y_2(t)}$$

$$\Rightarrow U_1 = \sum_{j=1}^n \delta_j W_1(T_j) \sum_{h=1}^2 R_{2,h}(T_j) Z_{j,h-1}(T_j) =$$

$$\sum_{j=1}^n \delta_j W_1(T_j) [R_{2,1}(T_j) + R_{2,2}Z_{j,1}(T_j)] =$$

$$\sum_{j=1}^n \delta_j \left( \frac{1}{Y_1(t)} + \frac{1}{Y_2(t)} \right) \left[ -\frac{1}{Y_2(T_j)} + Z_{j,1}(T_j) \left( \frac{1}{Y_1(T_j)} + \frac{1}{Y_2(T_j)} \right) \right] =$$

$$\sum_{j=1}^n \delta_j Z_{j,1}(T_j) - \sum_{j=1}^n \delta_j \left( \frac{Y_1(T_j)}{Y_1(T_j)+Y_2(T_j)} \right)$$

Damit entspricht die Teststatistik  $U_1$  der Differenz zwischen der Anzahl eingetretener Ereignisse und der Anzahl erwarteter Ereignisse in der ersten Stichprobe.

$$\begin{aligned}
\sigma_{1,1}^2 &= \sum_{j=1}^n \delta_j W_1(T_j)^2 \sum_{n=1}^2 \sum_{m=1}^2 R_{2,m}(T_j) R_{2,n}(T_j) Z_{j,m-1}(T_j) Z_{j,n-1}(T_j) = \\
&\sum_{j=1}^n \delta_j R_{2,2}^2 [R_{2,1}(T_j)^2 + R_{2,1}(T_j) R_{2,2}(T_j) Z_{j,1}(T_j) + R_{2,2}(T_j) R_{2,1}(T_j) Z_{j,1}(T_j)] = \\
&\sum_{j=1}^n \delta_j \left( \frac{Y_1(T_j) Y_2(T_j)}{Y_1(T_j) + Y_2(T_j)} \right)^2 \left[ \frac{1}{Y_2(T_j)^2} + Z_{j,1}(T_j) \left[ \frac{-2}{Y_2(T_j)} \left( \frac{1}{Y_1(T_j)} + \frac{1}{Y_2(T_j)} \right) + \left( \frac{1}{Y_1(T_j)} + \frac{1}{Y_2(T_j)} \right)^2 \right] \right] = \\
&\sum_{j=1}^n \delta_j \frac{Y_1(T_j)^2}{(Y_1(T_j) + Y_2(T_j))^2} - 2 \sum_{j=1}^n \frac{Y_1(T_j)}{Y_1(T_j) + Y_2(T_j)} Z_{j,1} + \sum_{j=1}^n \delta_j Z_{j,1} = \\
&\sum_{Stichprobe2} \delta_j \frac{Y_1(T_j)^2}{(Y_1(T_j) + Y_2(T_j))^2} + \sum_{Stichprobe1} \delta_j \frac{Y_2(T_j)^2}{(Y_1(T_j) + Y_2(T_j))^2}
\end{aligned}$$

Damit gilt:

$$\frac{U_1}{\sigma_{1,1}} \sim N(0, 1)$$

### 2.3 Beispiel

Der 2 Stichprobentest soll auf das Brustkrebspatientinnen Beispiel angewandt werden, wobei Stichprobe 1 immunoperoxidase-positive Patientinnen und Stichprobe 2 immunoperoxidase-negative Patientinnen sein sollen. Man erhält

$$U_1 = -4.187 \text{ und } \sigma_{1,1} = 6.04$$

$$\Rightarrow \frac{U_1}{\sqrt{\sigma_{1,1}}} = \frac{-4.187}{\sqrt{6.04}} = -1.704$$

Der p-Wert der Standardnormalverteilung ist 0.088. Damit gibt es zum Signifikanzniveau von 5% keinen Unterschied zwischen den beiden Gruppen.

Wird der Lograng Test benutzt, so erhält man die Teststatistik

$$\frac{-4.187}{\sqrt{3.1912}} = -2.344 \text{ mit einem p-Wert von } 0.019$$

Der Unterschied in den Tests rührt von den unterschiedlichen Schätzungen der Varianz von  $U_1$  her.

### 3 Kontraste

Bei einigen Problemen kann es interessant sein, Kontraste zu betrachten und auch zu testen, ob sie gleich null sind.

$$\text{Kontrast: } \sum_{i=0}^n \gamma_i \beta_i$$

#### 3.1 Beispiel

$$\gamma = (0, 1, 0, -1, 0, \dots, 0) \Rightarrow \sum_{i=0}^n \gamma_i \beta_i = \beta_1 - \beta_3$$

Unter  $H_0$  gilt somit  $\beta_1 - \beta_3 = 0$

#### 3.2 Teststatistik

Für  $r$  Kontraste wird eine  $r \times (p+1)$  Kontrastmatrix gebildet

$$\mathbf{C} = \begin{pmatrix} \gamma_{1,1} & \gamma_{1,2} & \dots & \gamma_{1,p+1} \\ \vdots & & & \\ \vdots & & & \\ \gamma_{r,1} & \gamma_{r,2} & \dots & \gamma_{r,p+1} \end{pmatrix} = \begin{pmatrix} C_1 \\ C_2 \\ \vdots \\ C_r \end{pmatrix}$$

Die Hypothese  $H_0$  ist damit

$$H_0 = \mathbf{C}\beta = \begin{pmatrix} C_1 \\ C_2 \\ \vdots \\ C_r \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Um nun einen Test zu konstruieren, wird eine modifizierte Gewichtsfunktion verwendet. Dazu sollen die Matrix  $\mathbf{Y}$  und der Vektor  $\mathbf{D}$  definiert werden.

$$\mathbf{Y}(T_i) = \begin{pmatrix} Y_1(T_i) & Y_1(T_i)Z_{11}(T_i) & \dots & Y_1(T_i)Z_{1p}(T_i) \\ \vdots & & & \\ \vdots & & & \\ Y_n(T_i) & Y_n(T_i)Z_{n1}(T_i) & \dots & Y_n(T_i)Z_{np}(T_i) \end{pmatrix}$$

Sind  $t_1 < t_2 < \dots < t_D$   $D$  unterschiedliche Ereigniszeiten (Todeszeitpunkte), so soll  $\mathbf{D}(t_i)$  ein  $n \times 1$  Vektor mit einer 1 in der  $j$ -ten Zeile genau dann wenn, Beobachtung  $j$  zum Zeitpunkt  $t_i$  das Ereignis erfährt, und 0 sonst.

Die modifizierte Gewichtsfunktion  $\mathbf{W}_c(\mathbf{t})$  sei

$$\mathbf{W}_c(t) = [\text{diag}\{\mathbf{C}[\mathbf{Y}^t(t)\mathbf{Y}(t)]^{-1}\mathbf{C}^t\}]^{-1} = [\text{diag}\{\mathbf{C}\mathbf{R}(t)\mathbf{C}^t\}]^{-1}$$

Damit ist die Teststatistik  $\mathbf{U}$  für  $r$  Kontraste

$$\mathbf{U}_c = \sum_{i=1}^D \mathbf{W}_c(t_i)\mathbf{C}\mathbf{R}(t_i)\mathbf{Y}(T_i)^t\mathbf{D}(t_i)$$

Und der Schätzer der Varianz-Kovarianz-Matrix ist

$$\mathbf{V}_c = \sum_{i=1}^D \mathbf{W}_c(t_i) \mathbf{C} \mathbf{R}(t_i) \mathbf{Y}(t_i)^t \text{diag}\{\mathbf{D}(t_i)\} \mathbf{Y}(T_i) \mathbf{R}^t(T_i) \mathbf{C}^t \mathbf{W}_c(T_i)^t$$

Unter  $H_0$  gilt asymptotisch

$$\mathbf{U}_c^t \mathbf{V}_c^{-1} \mathbf{U}_c \stackrel{as}{\sim} \chi_r^2$$

### 3.3 Beispiel

Die Methode der Kontraste soll auf die vier Stufen des Kehlkopfkrebsses angewandt werden. Konkret soll getestet werden ob Unterschiede in der Überlebenswahrscheinlichkeit zwischen den Stufen bestehen. Die Null-Hypothese soll sein, dass es, kontrolliert man für das Alter der Patienten, keine Unterschiede zwischen den Stufen gibt.

Zu Beginn wird mit den Methoden aus Sektion 2 überprüft, ob die Stufen als Ganzes betrachtet überhaupt einen Einfluss auf die Überlebenswahrscheinlichkeit haben.

$$\mathbf{U}_q = \begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix} = \begin{pmatrix} 0.707 \\ 4.284 \\ 7.812 \end{pmatrix}$$

$$\mathbf{V}_q = \begin{pmatrix} 3.437 & 1.309 & 0.453 \\ 1.309 & 6.105 & 0.449 \\ 0.453 & 0.449 & 7.210 \end{pmatrix}$$

Damit ist die Teststatistik

$$\mathbf{U}_q^t \mathbf{V}_q^{-1} \mathbf{U}_q = 10.947$$

Dies entspricht einem P-Wert von 0.012 einer  $\chi_3^2$ .

Auf einem Signifikanzniveau von 5% wird die Null-Hypothese verworfen. Mindestens eine Stufe hat einen von Null verschiedenen Einfluss. Nun, dass das Vorhandensein von Krebs einen Einfluss auf die Lebensdauer hat, ist ein relativ intuitives Resultat. Weniger einfach zu beantworten ist die Frage, ob einfaches Vorhandensein genügt, oder ob die unterschiedlichen Stufen auf verschiedene Weise Einfluss ausüben.

Zuerst soll mit Kontrasten getestet werden, ob ein Unterschied zwischen der ersten Stufe, Vorhandensein von Kehlkopfkrebs, und jeweils einer der restlichen 3 Stufen, die fortgeschrittene Krebsstadien beschreiben, besteht. Die Kontrastvektoren sind jeweils

$$\begin{aligned} c_1 &= (1 \quad -1 \quad 0 \quad 0 \quad 0) \\ c_2 &= (1 \quad 0 \quad -1 \quad 0 \quad 0) \\ c_3 &= (1 \quad 0 \quad 0 \quad -1 \quad 0) \\ c_4 &= (1 \quad 0 \quad 0 \quad 0 \quad -1) \end{aligned}$$

Effekt	Chi-Quadrat	d.f	p-Wert
$Z_1$ : Stufe 2	0.1456	1	0.7027
$Z_2$ : Stufe 3	3.0062	1	0.0829
$Z_3$ : Stufe 4	8.4655	1	0.0036
$Z_4$ : Alter	0.2333	1	0.6291

Auf einem Signifikanzniveau von 5% scheinen nur Patienten im 4. Stadion eine andere Überlebensrate zu haben als solche im ersten.

Desweiteren soll getestet werden, ob sich die Stufen 2, 3, 4 in ihrem Einfluss voneinander unterscheiden. Dazu sollen gleichzeitig folgende Hypothesen getestet werden:

$$\begin{aligned} \beta_1(t) &= \beta_2(t) \\ \beta_2(t) &= \beta_3(t) \end{aligned}$$

$$\Rightarrow H_0 : \beta_1(t) = \beta_2(t) = \beta_3(t)$$

Mit der Kontrastmatrix  $\mathbf{C}$  und p-Werten aus einer Chi-Quadrat-Verteilung mit 2 Freiheitsgraden folgt

$$\mathbf{C} = \begin{pmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \end{pmatrix}$$

$$\mathbf{U}_q = 6.813 \Rightarrow p_{\chi^2} = 0.0332 < 5\%$$

Die Null-Hypothese wird verworfen, folglich hat bestimmt eine der drei Stufen eine unterschiedliche Überlebensrate als die anderen zwei. Es lohnt sich demnach noch die drei Stufen einzeln gegeneinander zu testen.

Hypothese	Kontrastvektor	Teststatistik	p-Werte
$\beta_1(t) = \beta_2(t)$	$c = (0,1,-1,0,0)$	1.5884	0.2076
$\beta_1(t) = \beta_3(t)$	$c = (0,1,0,-1,0)$	6.9390	0.0084
$\beta_2(t) = \beta_3(t)$	$c = (0,0,1,-1,0)$	3.4238	0.0643

Die Tests zeigen, dass die Überlebensrate von Patienten im zweiten Stadion von solchen im vierten Stadion verschieden ist.