

Das additive Hazard Modell

Sandro Giammarino

25th June 2006

Beim *Proportional Hazard Modell* hatten wir gesehen, dass die Kovariablen multiplikativ auf einer unbekanntem baseline Hazard Funktion wirken. Die Schätzung der Risiko-Koeffizienten basierte auf dem *partial likelihood*. Im Proportional Hazards Modell waren diese Risiko-Koeffizienten zeitunabhängige, unbekannte Konstanten.

In diesem Kapitel präsentieren wir ein alternatives Modell, basierend darauf, dass die Kovariablen additiv auf eine unbekanntem baseline Hazard Rate wirken. Die unbekanntem Risiko-Koeffizienten in diesem Modell dürfen zeitabhängige Funktionen sein, d.h. die Effekte der Kovariablen können mit der Zeit variieren. Im Proportional Hazards Modell hingegen benutzte man likelihood basierende Schätzungstechniken; die Schätzer der Risiko-Koeffizienten dieses Modells basieren auf der Methode der kleinsten Quadrate. Die Herleitung basiert auf der Annäherung eines Zählprozesses der Überlebensanalyse, welche ähnlich zu der Herleitung des Nelson-Aalen Schätzers der kumulativen Hazard Rate ist. Die Hypothesentests basieren auf stochastischen Integralen der sich ergebenden Schätzer. In diesem Kapitel werden wir eine nicht gewichtete Schätzung der kleinsten Quadrate präsentieren.

Die Daten bestehen aus einer Stichprobe

$$[T_j, \delta_j, \mathbf{Z}_j(t)], j = 1, \dots, n$$

wobei T_j die Dauer der Studie, δ_j der Ereignisindikator und $\mathbf{Z}_j(t) = [Z_{j1}(t), \dots, Z_{jp}(t)]$ ist ein p -Vektor mit, möglicherweise, zeitabhängigen Kovariablen. Für das j te Individuum definieren wir den stochastischen Prozess

$$Y_j(t) = \begin{cases} 1 & \text{falls das } j\text{te Individuum unter Beobachtung (d.h. unter Risiko) zur Zeit } t \text{ ist} \\ 0 & \text{falls das } j\text{te Individuum nicht unter Beobachtung ist} \end{cases}$$

Wenn die Daten linksseitig gestutzt sind, gilt $Y_j(t) = 1$ nur zwischen der Eintrittszeit und Endzeit der Studie eines Individuums. Für rechtsseitig zensurierte Daten gilt $Y_j(t) = 1$ wenn $t \leq T_j$.

Für das j te Individuum modellieren wir die *bedingte Hazard Rate* zur Zeit t mit gegebenen $\mathbf{Z}_j(t)$ durch

$$h [t|\mathbf{Z}_j(t)] = \beta_0(t) + \sum_{k=1}^p \beta_k(t)Z_{jk}(t)$$

wobei die unbekannt Parameterfunktionen $\beta_k(t)$ für $k = 1, \dots, p$ zu schätzen sind. Direkte Schätzung dieser $\beta_k(t)$ ist sehr schwierig in der Praxis. Wir schätzen die *kumulative Risiko Funktion* $B_k(t)$, definiert durch

$$B_k(t) := \int_0^t \beta_k(u)du, k = 0, 1, \dots, p.$$

Die grobe Schätzung von $\beta_k(t)$ ist gegeben durch die Steigung unserer Schätzung von $B_k(t)$. Bessere Schätzungen von $\beta_k(t)$ erhält man mit *kernel-smoothing Techniken*. Diese Methoden werden wir aber nicht genauer behandeln. Eine Anwendung dafür gibt es im zweiten Beispiel weiter unten.

Herleitung der Schätzer

Wie schon erwähnt, basieren die Schätzer $B_k(t)$ auf der Methode der kleinsten Quadrate. Für die Herleitung definieren wir für jedes Individuum einen Zählprozess

$$N_j(t) := I_{\{T_j \leq t, \delta_j = 1\}}$$

Sei nun $\mathbf{N}(t) = [N_1(t), \dots, N_n(t)]^T$ der Spaltenvektor von n Zählprozessen und $\mathbf{Y}(t)$ eine $n \times (p+1)$ Matrix mit Zeileneinträgen

$$[Y_j(t), Y_j(t)Z_{j1}(t), \dots, Y_j(t)Z_{jp}(t)], j = 1, \dots, n$$

Multiplizieren wir $\mathbf{Y}(t)$ mit dem Spaltenvektor $\beta(t) = [\beta_0(t), \dots, \beta_p(t)]^T$, so erhalten wir mit

$$\mathbf{M}(t) = \mathbf{N}(t) - \int_0^t \mathbf{Y}(u)\beta(u)du$$

einen Spaltenvektor von n Martingalen. daraus folgt

$$d\mathbf{N}(t) = \mathbf{Y}(t)\beta(t) + d\mathbf{M}(t)$$

Die Martingale können nun als statistische Störungen angesehen werden, deshalb setzen wir $d\mathbf{M}(t) = 0$, um den Fehler zu minimieren. Daraus können wir nun die Gleichung $d\mathbf{N}(t) = \mathbf{Y}(t)\beta(t)$ lösen, um den Vektor $\mathbf{B}(t) = \int_0^t \beta(u)du$ zu erhalten. Eine Lösung ist nur möglich, wenn $\mathbf{Y}(t)$ den maximalen Rang hat, nämlich $p+1$. Wir schreiben nun $\mathbf{Y}^-(t)$ für die allgemeine Inverse von $\mathbf{Y}(t)$, d.h. $\mathbf{Y}^-(t)$ ist dann eine $(p+1) \times n$ Matrix mit $\mathbf{Y}^-(t)\mathbf{Y}(t) = \mathbf{Id}_{(p+1)}$. Die Lösung der Gleichung $d\mathbf{N}(t) = \mathbf{Y}(t)\beta(t)$ ist dann der geschätzte Wert

$$\widehat{\mathbf{B}}(t) = \int_0^t \mathbf{Y}^-(u) d\mathbf{N}(u) = \sum_{T_j \leq t} \delta_j \mathbf{Y}^-(T_j)$$

mit $t \leq \tau := \max\{T_j\}$. Hier haben wir die allgemeine Inverse

$$\mathbf{Y}^-(t) = [\mathbf{Y}^T(t)\mathbf{Y}(t)]^{-1} \mathbf{Y}^T(t)$$

benutzt.

Um die Schätzungen zu erhalten, berechnen wir zuerst die $(n+1) \times (n+1)$ Matrix $\mathbf{A}(t)$, dessen (g, k) te Element gegeben ist durch

$$a_{g,k}(t) = \sum_{j=1}^n Y_j(t) Z_{j(g-1)}(t) Z_{j(k-1)}(t)$$

für $g, k = 1, \dots, p+1$, wobei $Z_{j0}(t) = 1$.

Sei $\mathbf{A}(T_j)$ eine nicht singuläre Matrix. Wir definieren $R(t) := \mathbf{A}^{-1}(t)$, bestehend aus den Elementen $R_{g,k}(t)$. Der Schätzer von $B_k(t)$ ist dann gegeben durch

$$\widehat{B}_g(t) = \sum_{T_j \leq t} \delta_j \sum_{k=1}^{p+1} R_{g+1,k}(T_j) Z_{j(k-1)}(T_j)$$

für $g = 0, \dots, p$ und $t \leq \tau$.

Herleitung der Varianz der Schätzer

Um den Schätzer für die Varianz von $\widehat{B}(t)$ mit $t \leq \tau$ zu erhalten, schreiben wir das stochastische Integral

$$(\widehat{\mathbf{B}} - \mathbf{B})(t) = \int_0^t \mathbf{Y}^-(u) d\mathbf{M}(u)$$

des previsible Prozesses $\mathbf{Y}^-(t)$ bezogen auf ein Martingal. Der previsible Variationsprozess $\langle \widehat{\mathbf{B}} - \mathbf{B} \rangle(t)$ gibt den mittleren quadratischen Fehler für $(\widehat{\mathbf{B}} - \mathbf{B})(t)$, gegeben durch

$$\langle \widehat{\mathbf{B}} - \mathbf{B} \rangle(t) = \int_0^t \mathbf{Y}^-(u) \langle d\mathbf{M}(u) \rangle \mathbf{Y}^-(u)^T = \int_0^t \mathbf{Y}^-(u) [\text{diag } \mathbf{h}(u)] \mathbf{Y}(u)^T$$

wobei $h[u|\mathbf{Z}_j(u)] Y_j(u)$ das (j, j) te Element der Diagonalmatrix $\text{diag } \mathbf{h}(u)$ ist. Die geschätzte Varianz erhält man mit der Matrix

$$\int_0^t \mathbf{Y}^-(u) [\text{diag } d\mathbf{N}(u)] \mathbf{Y}(u)^T$$

wobei $\text{diag } d\mathbf{N}(u)$ die Diagonalelemente $dN_j(u)$ enthält.

Der Schätzer für die Kovarianz zwischen $\widehat{B}_g(t)$ und $\widehat{B}_k(t)$ ist gegeben durch

$$\widehat{Cov} \left[\widehat{B}_g(t), \widehat{B}_k(t) \right] = \sum_{T_j \leq t} \delta_j \sum_{m=1}^{p+1} \sum_{h=1}^{p+1} R_{g+1,m}(T_j) R_{k+1,h}(T_j) Z_{j(m-1)}(T_j) Z_{j(h-1)}(T_j)$$

für $g, k = 0, \dots, p$. Die Schätzer $\widehat{B}_k(t)$ schätzen das Integral der Regressionsfunktionen β_k in gleicher Weise wie der Nelson-Aalen Schätzer; sie schätzen das Integral der Hazard Rate im univariaten Fall. Konfidenzintervalle für die integrierten Regressionsfunktionen $\widehat{B}_k(t)$, basierend auf der linearen Formulierung, sind in genau gleicher Weise konstruiert wie die Konfidenzintervalle für die kumulative Hazard Funktionen. Hier werden einfach der Nelson-Aalen Schätzer $\widehat{\Lambda}(t)$ und $\sigma_\Lambda(t)$ in $\widehat{\Lambda}(t) \pm Z_{1-\frac{\alpha}{2}}(t) \sigma_\Lambda(t)$ mit $\widehat{B}_k(t)$ und $\widehat{Var} \left[\widehat{B}_k(t) \right]^{1/2}$ ersetzt.

Schätzer und Varianz im Fall $n=2$

Um diese Schätzer zu illustrieren, nehmen wir ein Zweistichprobenproblem, d.h. $j = 1, 2$. In diesem Fall haben wir für eine einzelne Kovariable

$$Z_{j1} = \begin{cases} 1 & \text{falls die } j\text{te Beobachtung von der Stichprobe 1 ist} \\ 0 & \text{falls die } j\text{te Beobachtung in der Stichprobe 2 ist} \end{cases}$$

Dann ist

$$\mathbf{A}(t) = \begin{pmatrix} Y_1(t) + Y_2(t) & Y_1(t) \\ Y_1(t) & Y_1(t) \end{pmatrix}$$

Die Matrix $R(t)$ ist dann gegeben durch

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

wobei $Y_k(t)$ die Anzahl unter Risiko in der Stichprobe $k = 1, 2$ ist zur Zeit t . Setzen wir diese Matrix in $\widehat{B}_g(t)$ ein, so erhalten wir:

$$\widehat{B}_0(t) = \sum_{T_j \leq t} \frac{\delta_j}{Y_2(T_j)} - \sum_{\substack{T_j \leq t \\ j \in \text{Stichprobe 1}}} \frac{\delta_j}{Y_2(T_j)} = \sum_{\substack{T_j \leq t \\ j \in \text{Stichprobe 2}}} \frac{\delta_j}{Y_2(T_j)} \quad (1)$$

für $t \leq \tau$. Dies ist der Nelson-Aalen Schätzer der Hazard Rate, welcher nur die Daten der zweiten Stichprobe benutzt. Analog

$$\begin{aligned}
\widehat{B}_1(t) &= -\sum_{T_j \leq t} \frac{\delta_j}{Y_2(T_j)} + \sum_{\substack{T_j \leq t \\ j \in \text{Stichprobe1}}} \delta_j \left(\frac{1}{Y_1(T_j)} + \frac{1}{Y_2(T_j)} \right) \\
&= \sum_{\substack{T_j \leq t \\ j \in \text{Stichprobe1}}} \frac{\delta_j}{Y_1(T_j)} - \sum_{\substack{T_j \leq t \\ j \in \text{Stichprobe2}}} \frac{\delta_j}{Y_2(T_j)}
\end{aligned} \tag{2}$$

für $t \leq \tau$. Dies ist der Unterschied zwischen den Nelson-Aalen Schätzer der Hazard Rate in den beiden Stichproben. Die Schätzung der Varianz und Kovarianz Funktionen von $\widehat{B}_0(t)$ und $\widehat{B}_1(t)$ sind gegeben durch

$$\begin{aligned}
\widehat{Var} \left[\widehat{B}_0(t) \right] &= \sum_{T_j \leq t} \frac{\delta_j}{Y_2(T_j)^2} - \sum_{\substack{T_j \leq t \\ j \in \text{Stichprobe1}}} \frac{\delta_j}{Y_2(T_j)^2} \\
&= \sum_{\substack{T_j \leq t \\ j \in \text{Stichprobe2}}} \frac{\delta_j}{Y_2(T_j)^2}
\end{aligned} \tag{3}$$

$$\begin{aligned}
\widehat{Var} \left[\widehat{B}_1(t) \right] &= \sum_{T_j \leq t} \frac{\delta_j}{Y_2(T_j)^2} \\
&+ \sum_{\substack{T_j \leq t \\ j \in \text{Stichprobe1}}} \delta_j \left(\frac{-2}{Y_2(T_j)} \left[\frac{1}{Y_1(T_j)} + \frac{1}{Y_2(T_j)} \right] + \left[\frac{1}{Y_1(T_j)} + \frac{1}{Y_2(T_j)} \right]^2 \right) \\
&= \sum_{\substack{T_j \leq t \\ j \in \text{Stichprobe1}}} \frac{\delta_j}{Y_1(T_j)^2} + \sum_{\substack{T_j \leq t \\ j \in \text{Stichprobe2}}} \frac{\delta_j}{Y_2(T_j)^2}
\end{aligned} \tag{4}$$

und

$$\begin{aligned}
\widehat{Cov} \left[\widehat{B}_0(t), \widehat{B}_1(t) \right] &= \sum_{T_j \leq t} \frac{-\delta_j}{Y_2(T_j)^2} + \sum_{\substack{T_j \leq t \\ j \in \text{Stichprobe1}}} \frac{\delta_j}{Y_2(T_j)^2} \\
&= -\sum_{\substack{T_j \leq t \\ j \in \text{Stichprobe2}}} \frac{\delta_j}{Y_2(T_j)^2}, \quad t \leq \tau.
\end{aligned} \tag{5}$$

Beispiel zum additiven Hazard Modell

Als ein Beispiel für das additive Modell vergleichen wir die vier Stadien eines Kehlkopfkrebsses, in der Studie auf 90 männliche Kehlkopfkrebs-Patienten angepasst. Wir definieren vier Kovariablen zu einer bestimmten Zeit:

$$Z_1 = \begin{cases} 1 & \text{falls in Stadium II der Krankheit} \\ 0 & \text{falls in Stadium I,III oder IV der Krankheit;} \end{cases}$$

$$Z_2 = \begin{cases} 1 & \text{falls in Stadium III der Krankheit} \\ 0 & \text{falls in Stadium I,II oder IV der Krankheit;} \end{cases}$$

$$Z_3 = \begin{cases} 1 & \text{falls in Stadium IV der Krankheit} \\ 0 & \text{falls in Stadium I,II oder III der Krankheit;} \end{cases}$$

$$Z_4 = \text{Alter bei Diagnose} - 64.11$$

Hier haben wir Z_4 bei seinem Erwartungswert zentriert. In diesem Beispiel haben wir $\tau = 4.4$, wenn mindestens ein Patient immer noch unter Risiko jeder dieser vier Krankheitsstadien ist.

die Grafiken 10.3-10.7 zeigen die Schätzungen der $\hat{B}_k(t)$, $k = 0, 1, \dots, 4$ mit 95%punktweise Konfidenzintervalle. In der Grafik 10.3 ist die geschätzte kumulative baseline Hazard $\hat{B}_0(t)$ eine Schätzung der kumulative Hazard Rate der Stufe I Patienten im Alter von 64.11 Jahren. $\hat{B}_1(t)$, $\hat{B}_2(t)$ und $\hat{B}_3(t)$ (Grafiken 10.4-10.6) zeigen das kumulative Exzessrisiko der Patienten mit ähnlichen Alter in den Stufen II,III und IV. Hier scheint es, dass es ein kleines Risiko gibt für Patienten der Stufe II. Patienten der Stufe III und IV haben erhöhtes Risiko in den zwei ersten Jahren, wenn die Steigungen der Diagnosen der beiden kumulativen Hazards nicht Null sind. Graphik 10.7 zeigt das Exzessrisiko zum Alter.

Um die Unterschiede der Überlebenschancen in den verschiedenen Stufen der Krankheiten zu verstehen, betrachten wir Grafik 10.8, die eine kernel-geglättete Schätzung von $\beta_k(t)$ mit $k = 1, 2, 3$ ist. (In diesem Fall wurde eine biweight kernel benutzt über einem Jahr). Diese Grafik zeigt, dass es ein kleines Exzessrisiko in der Stufe II Krankheit gibt, dass das Exzessrisiko in der Stufe III Krankheit negiert wird nach zweieinhalb Jahren, und dass die Stufe IV Krankheit der schlechteste Fall mit Exzessrisiko etwa bei 0.4 Tote pro Jahr während den ersten zwei Jahren ist nach Transplantation.

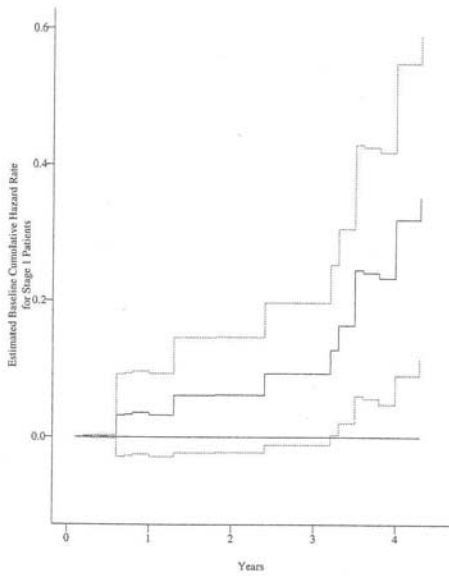


Figure 10.3 Estimate of the cumulative baseline hazard rate (B_0) and a 95% pointwise confidence interval for laryngeal cancer patients.

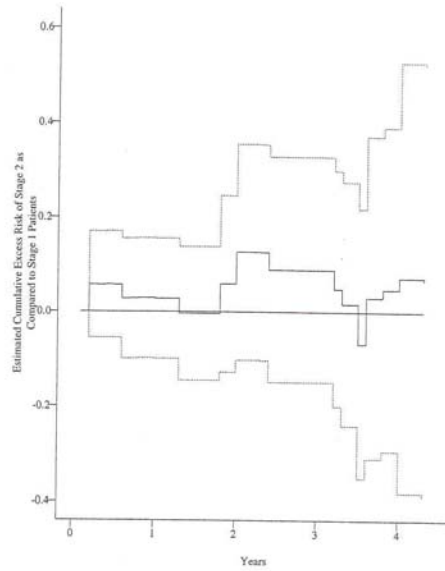


Figure 10.4 Estimate of the cumulative excess risk of Stage II cancer as compared to Stage I cancer (B_1) and a 95% pointwise confidence interval for laryngeal cancer patients.

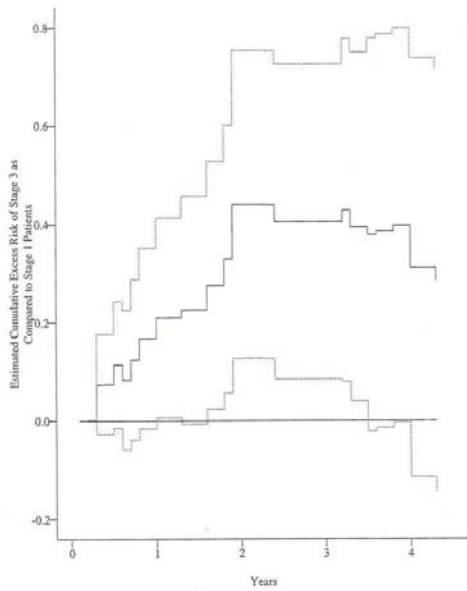


Figure 10.5 Estimate of the cumulative excess risk of Stage III cancer as compared to Stage I cancer (B_2) and a 95% pointwise confidence interval for laryngeal cancer patients.

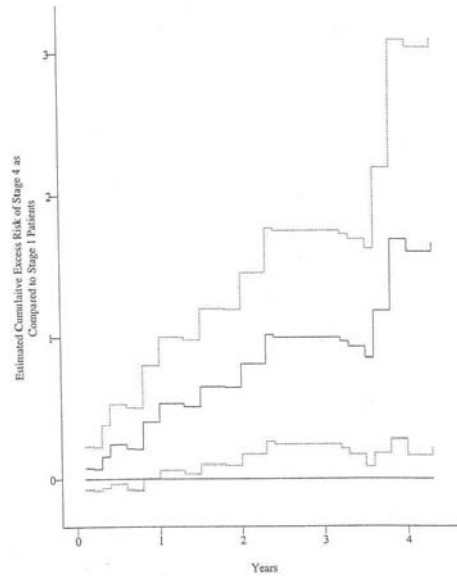


Figure 10.6 Estimate of the cumulative excess risk of Stage IV cancer as compared to Stage I cancer (B_3) and a 95% pointwise confidence interval for laryngeal cancer patients.

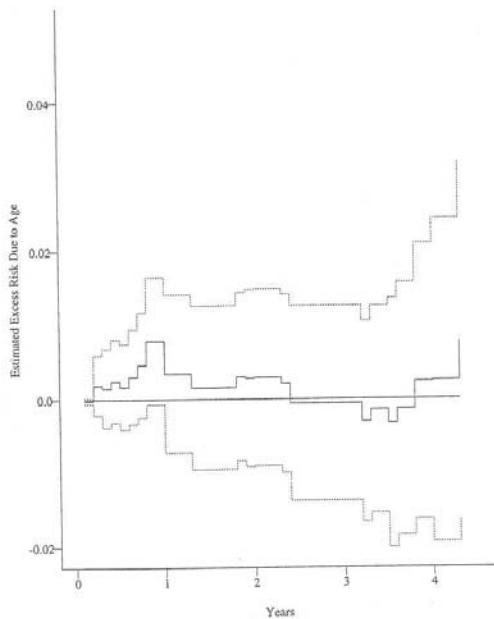


Figure 10.7 Estimate of the cumulative effect of age (B_4) and a 95% pointwise confidence interval for laryngeal cancer patients.

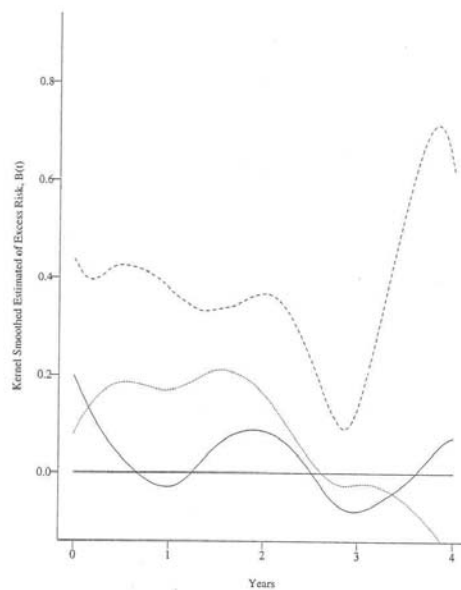


Figure 10.8 Kernel smoothed estimates of the excess risk of death for Stage II, III, or IV cancer as compared to Stage I cancer. Stage II (—) Stage III (---) Stage IV (·····).