Bayesian Model Averaging im Survival Kontext

Alexander Volkmann

20. Juni 2017

Seminar Moderne statistische Methoden in der Epidemiologie
Ludwig-Maximilians-Universität München

Gliederung

Gliederung

Primäre Biliäre Zirrhose

Datengrundlage

Eine von 1974 bis 1984 durchgeführte randomisierte klinische Studie an 312 Patienten mit Primärer Biliärer Zirrhose

Ziel

Modell für Verlauf der Krankheit erstellen
(Welche Variablen haben Einfluss auf das Versterben der Patienten?
Prognose für einzelne Patienten?)

Primäre Biliäre Zirrhose

Datengrundlage
Eine von 1974 bis 1984 durchgeführte randomisierte klinische Studie an 312 Patienten mit Primärer Biliärer Zirrhose

Ziel
Modell für Verlauf der Krankheit erstellen
(Welche Variablen haben Einfluss auf das Versterben der Patienten? Prognose für einzelne Patienten?)

`pbc`-**Datensatz**

- Anzahl der Tage zwischen Registrierung und dem Eintreten der Transplantation, des Todes oder des Studienendes
- Status am Enpunkt (zensiert, transplantiert oder tot)
- 15 mögliche Risikofaktoren (Kovariablen)

pbc-Datensatz

Variable	Variablenbeschreibung	Mittelwert
age	Alter in Jahre	50,02
albumin	Albumin im Serum (g/dl) [logarithmiert]	1,25
bili	Bilirubin im Serum (mg/dl) [logarithmiert]	0,58
edema	Ödem (0: nicht vorhanden; 0,5: unbehandelt oder erfogreich behandelt; 1: Ödem trotz Therapie)	0,11
protime	Standardisierte Blutgerinnungszeit [logarithmiert]	2,37
stage	Histologisches Krankheitsstadium, Biopsie benötigt (1-4)	3,03
copper	Kupfer im Urin (ug/Tag) [logarithmiert]	4,26
$\vdots$	$\vdots$	$\vdots$

Modellierung der Überlebenszeit

Lebensdauer-Modelle

Cox Proportional Hazards Modell

$$\lambda(t \mid X_i) = \lambda_0(t)exp(X_i\beta)$$

t	Zeitpunkt
X_i	Vektor der Kovariablenausprägungen für Patient i
β	Vektor der unbekannten Parameter
λ	Hazard-Rate
λ_0	Baseline-Hazard (unspezifiziert)

„Standard“-Vorgehen

Backwards-Elimination

Beliebte Variablenselektionsmethode:

1. Beginne mit vollem Modell
2. Berechne p-Werte der einzelnen Koeffizienten
3. Entferne die Variable aus dem Modell, deren p-Wert am größten ist
4. Passe Modell neu an
5. Wiederhole Schritte 2 bis 4 bis alle p-Werte $< 0,05$

„Standard"-Vorgehen

Ergebnis des BE-Modells

	coef	exp(coef)	se(coef)	z	p
age	0.03	1.03	0.01	3.66	0.00
albumin	-2.83	0.06	0.74	-3.82	0.00
bili	0.76	2.14	0.11	6.63	0.00
edema	0.82	2.27	0.30	2.72	0.01
protime	3.06	21.39	1.10	2.78	0.01
copper	0.36	1.43	0.14	2.63	0.01

Probleme mit „Standard“-Vorgehen

Dieses Vorgehen ist problematisch!

1. Durch das wiederholte Testen verlieren die p-Werte ihre gewohnte Bedeutung

$\Rightarrow$ Signifikanzniveau für Inferenz nicht mehr bekannt

Probleme mit „Standard"-Vorgehen

Dieses Vorgehen ist problematisch!

1. Durch das wiederholte Testen verlieren die p-Werte ihre gewohnte Bedeutung

$\Rightarrow$ Signifikanzniveau für Inferenz nicht mehr bekannt

Lösungsansätze

- **Post-Selection Inference:**
  Z.B. Modellbildung und Inferenz auf unabhängigen Datensätzen durchführen

Probleme mit „Standard“-Vorgehen

2. Durch das Ausschließen von Variablen basierend auf statistische Tests wird auch das End-Modell „zufällig“

$\Rightarrow$ Auch ein anderes Modell hätte ausgewählt werden können

$\Rightarrow$ **Unsicherheit wird unterschätzt**

Probleme mit „Standard"-Vorgehen

2. Durch das Ausschließen von Variablen basierend auf statistische Tests wird auch das End-Modell „zufällig"

$\Rightarrow$ Auch ein anderes Modell hätte ausgewählt werden können

$\Rightarrow$ **Unsicherheit wird unterschätzt**

Lösungsansätze

- **Baysian Model Averaging:**
  Methode, um die Modellunsicherheit zu berücksichtigen
  $\Rightarrow$ Mittellung der Ergebnisse über mehrere Modelle

Gliederung

Abgrenzung zur klassischen Statistik
Parametervektor θ eines Wahrscheinlichkeitsmodells ist in der Bayesianischen Statistik zufällig und besitzt eine Verteilung

Abgrenzung zur klassischen Statistik
Parametervektor θ eines Wahrscheinlichkeitsmodells ist in der Bayesianischen Statistik zufällig und besitzt eine Verteilung

Anwendung des Satz von Bayes

$$f(\theta \mid x) = \frac{f(\theta, x)}{f(x)} = \frac{f(x \mid \theta) f(\theta)}{\int f(x \mid \theta) f(\theta) d\theta}$$

θ	Parametervektor
x	Realisierung der zufälligen Daten X
$f(\theta)$	Priori-Verteilung von θ
$f(x \mid \theta)$	Likelihood der Daten x
$f(\theta \mid x)$	Posteriori-Verteilung von θ

Bayes-Inferenz

Posteriori-Verteilung als Grundlage der Inferenz
Punktschätzer über Posteriori-Erwartungswert, -Median oder -Modus

Modellwahl
Über Marginale Likelihood

$$f(x \mid M_i) = \int f(x \mid \theta_i, M_i) f(\theta_i \mid M_i) d\theta_i$$

Bayes-Inferenz

Posteriori-Verteilung als Grundlage der Inferenz
Punktschätzer über Posteriori-Erwartungswert, -Median oder -Modus

Modellwahl
Über Marginale Likelihood

$$f(x \mid M_i) = \int f(x \mid \theta_i, M_i) f(\theta_i \mid M_i) d\theta_i$$

oder Bayes'sches Informationskriterium

$$BIC = 2 \log f(x \mid \hat{\theta}_{ML}, M_i) - p \log(n) \doteq 2 \log f(x \mid M_i)$$

mit Modell M_i, Maximum-Likelihood Schätzer $\hat{\theta}_{ML}$,p der Anzahl der Dimensionen von θ und n der Größe des Stichprobenumfangs

Gliederung

Grundidee

Berücksichtigung der Modelunsicherheit
Mehrere Modelle $M_1, ..., M_K$ werden an die Daten D angepasst

Mittelung über Modelle
Gewichtetes Mittel über die Modelle, um Aussagen über die interessierende Größe Δ zu treffen

$\Rightarrow$ Posteriori-Modellwahrscheinlichkeit als Gewichte

Posteriori-Modellwahrscheinlichkeit

Um die Wahrscheinlichkeit des Modells M_k gegeben die Daten D zu berechnen, kann wieder der Satz von Bayes verwendet werden:

$$P(M_k \mid D) = \frac{P(D \mid M_k)P(M_k)}{\sum_{l=1}^{K} P(D \mid M_l)P(M_l)},$$

wobei $P(D \mid M_k)$ bereits als marginale Likelihood bekannt ist:

$$P(D \mid M_k) = \int P(D \mid \theta_k, M_k)P(\theta_k \mid M_k)d\theta_k$$

und $P(M_k)$ die Priori-Modellwahrscheinlichkeit bezeichnet.

Prädiktive Verteilung von Δ

Verteilung von Δ in einem Modell
Um Aussagen über die interessierende Größe Δ zu treffen, wird in der Bayesianischen Statistik die Prädiktive (Posteriori-) Verteilung betrachtet:

$$P(\Delta \mid M_k, D) = \int P(\Delta \mid \theta_k, M_k, D) P(\theta_k \mid M_k, D) d\theta_k.$$

Posteriori-Verteilung von Δ

Berücksichtigung aller Modelle

Gewichtetes Mittel aus prädiktiver Verteilung und Posteriori-Modellwahrscheinlichkeit

$$P(\Delta \mid D) = \sum_{k=1}^{K} P(\Delta \mid M_k, D) P(M_k \mid D).$$

Posteriori-Verteilung von Δ

Berücksichtigung aller Modelle
Gewichtetes Mittel aus prädiktiver Verteilung und Posteriori-Modellwahrscheinlichkeit

$$P(\Delta \mid D) = \sum_{k=1}^{K} P(\Delta \mid M_k, D) P(M_k \mid D).$$

Um die Posteriori-Verteilung auszuwerten, kann beispielsweise der Posteriori-Erwartungswert berechnet werden

$$E[\Delta \mid D] = \sum_{k=1}^{K} E[\Delta \mid D, M_k] P(M_k | D),$$

Einfache Grundidee

Gewichtetes Mittel über mehrere Modelle

Einbettung in Bayesianischen Kontext

Annahme von Priori-Verteilungen, Untersuchung von Posteriori-Verteilungen

Bayesian Model Averaging

Einfache Grundidee
Gewichtetes Mittel über mehrere Modelle

Einbettung in Bayesianischen Kontext
Annahme von Priori-Verteilungen, Untersuchung von Posteriori-Verteilungen

Probleme in der Umsetzung

- Integral der prädiktiven Verteilung von Δ
- Integral der marginalen Likelihood
- Identifizierung der K Modelle

Gliederung

Beschränkung auf Cox-Modelle
Betrachte Modelle einer Modellklasse, die sich nur in den aufgenommenen Kovariablen unterscheiden

Überlebenszeitmodelle

Beschränkung auf Cox-Modelle
Betrachte Modelle einer Modellklasse, die sich nur in den aufgenommenen Kovariablen unterscheiden

Partial Likelihood
Besonderheit der Cox-Modelle

$$PL(\beta) = \prod_{i=1}^{m} \frac{exp(X_i^T \beta)}{\sum_{j \in R_i} exp(X_j^T \beta)}$$

$i = 1, ..., m$	Individuen mit Ereignis
$i = m+1, ..., n$	Zensierte Individuen
R_i	Risk-Set zum Zeitpunkt des Ereignisses für i

Approximation der prädiktiven Verteilung von Δ

Zur Erinnerung:

$$P(\Delta \mid M_k, D) = \int P(\Delta \mid \theta_k, M_k, D) P(\theta_k \mid M_k, D) d\theta_k.$$

Approximation der prädiktiven Verteilung von Δ

Zur Erinnerung:

$$P(\Delta \mid M_k, D) = \int P(\Delta \mid \theta_k, M_k, D) P(\theta_k \mid M_k, D) d\theta_k.$$

Maximum-Likelihood-Estimator Approximation

$$P(\Delta \mid M_k, D) \approx P(\Delta \mid M_k, D, \hat{\beta}_{k,ML})$$

Ursprünglich aus der robusten Likelihood-Schätzung für Zeitreihen

Approximation der marginalen Likelihood

Zur Erinnerung:

$$P(D \mid M_k) = \int P(D \mid \theta_k, M_k) P(\theta_k \mid M_k) d\theta_k$$

Approximation der marginalen Likelihood

Zur Erinnerung:

$$P(D \mid M_k) = \int P(D \mid \theta_k, M_k) P(\theta_k \mid M_k) d\theta_k$$

Approximation durch das BIC

$$\log(P(D \mid M_k)) \approx BIC.$$

Anstelle der Gesamtzahl der Beobachtungen n wird nur die Zahl der beobachteten Ereignisse m aufgenommen

$\Rightarrow$ Eher grobe Approximation (Theoretischer Fehlerterm $\mathcal{O}(1)$)

Modellmenge $\mathcal{M} = M_1, ..., M_K$

Über welche Modelle soll gemittelt werden?

- Mittelung über alle möglichen Kovariablenkombinationen

$\Rightarrow$ Bei p Variablen 2^p mögliche Modelle

$\Rightarrow$ Rechentechnisch sehr aufwendig

Modellmenge $\mathcal{M} = M_1, ..., M_K$

Über welche Modelle soll gemittelt werden?

- Mittelung über alle möglichen Kovariablenkombinationen

$\Rightarrow$ Bei p Variablen 2^p mögliche Modelle

$\Rightarrow$ Rechentechnisch sehr aufwendig

- Mittelung nur über Modelle, die von den Daten gestützt werden

$\Rightarrow$ Modelle mit den höchsten Posteriori-Modellwahrscheinlichkeiten

$$\mathcal{A} = \left\{ M_k : \frac{\max_l\{P(M_l \mid D)\}}{P(M_k \mid D)} \leq C \right\}$$

Leaps-and-Bounds Algorithmus

Schätzung der q besten Modelle für jede Modellgröße
Ohne Modelle tatsächlich zu berechnen werden die q Modelle mit der höchsten Posteriori-Modellwahrscheinlichkeit ausgegeben

$\Rightarrow$ Mehr Modelle identifiziert, als in $\mathcal{A}$ enthalten

Weitere Reduktionsschritte

- Approximativer Likelihood-Quotienten Test
- Berechnung der Modelle und Ausschluss über BIC

Gliederung

pbc-Datensatz

Zur Erinnerung:

- Überlebenszeit-Daten mit 15 Kovariablen
- Welche Risikofaktoren wirken auf das Überleben?
- BE-Modell mit 6 Variablen `age`, `albumin`, `bili`, `edema`, `protime`, `copper`

pbc-Datensatz

Zur Erinnerung:

- Überlebenszeit-Daten mit 15 Kovariablen
- Welche Risikofaktoren wirken auf das Überleben?
- BE-Modell mit 6 Variablen `age`, `albumin`, `bili`, `edema`, `protime`, `copper`

Anwendung von Bayesian Model Averaging

Leaps-and-Bounds Algorithmus ist in Funktion `bic.surv` im `R`-Package `BMA` implementiert

$\Rightarrow$ Aus $2^{15} = 32768$ möglichen Modellen werden 36 Modelle identifiziert

Bayesian Model Averaging Ergebnisse

Ausgewählte Ergebnisse

M_k	age	albumin	bili	edema	protime	copper	stage	PMW
M_1	0,03	-2,82	0,76	0,82	3,07	0,36	–	0,17
M_2	0,03	-2,47	0,75	0,82	2,67	0,35	0,24	0,07
M_3	0,03	-2,40	0,79	–	–	0,35	0,31	0.07
M_4	0,03	-3,28	0,80	–	3,74	0,35	–	0,06
M_5	0,03	-3,01	0,89	0,82	2,91	–	–	0,05
PEW	0,03	-2,78	0,79	0,74	2,43	0,25	0,10	
PEP	1	1	1	0,85	0,78	0,72	0,37	

PMW: Posteriori-Modellwahrscheinlichkeit, PEW: Posteriori-Erwartungswert, PEP: Posteriori-Effektwahrscheinlichkeit

Posteriori-Effektwahrscheinlichkeit

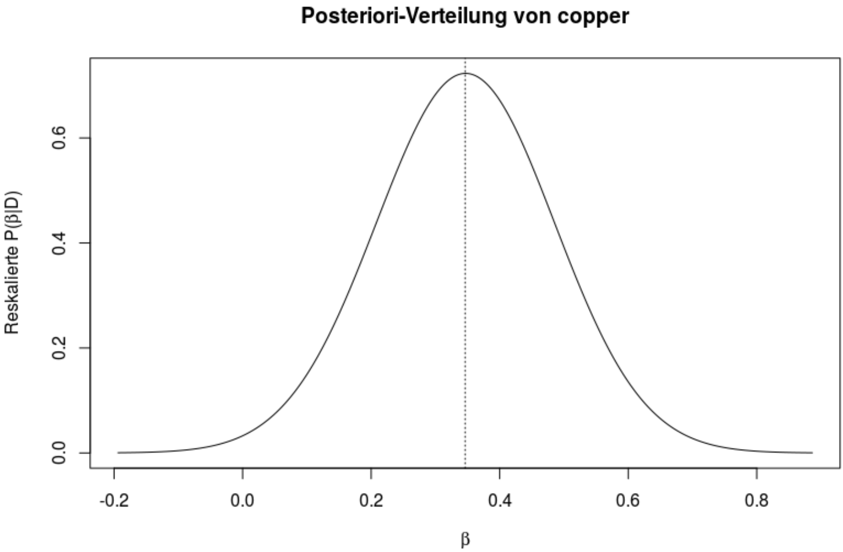
„Analogie“ zu p-Werten
$P(\beta \neq 0)$: Wahrscheinlichkeit, dass der jeweilige Koeffizient ungleich 0 ist

$\Rightarrow$ Interpretation einer Wahrscheinlichkeit

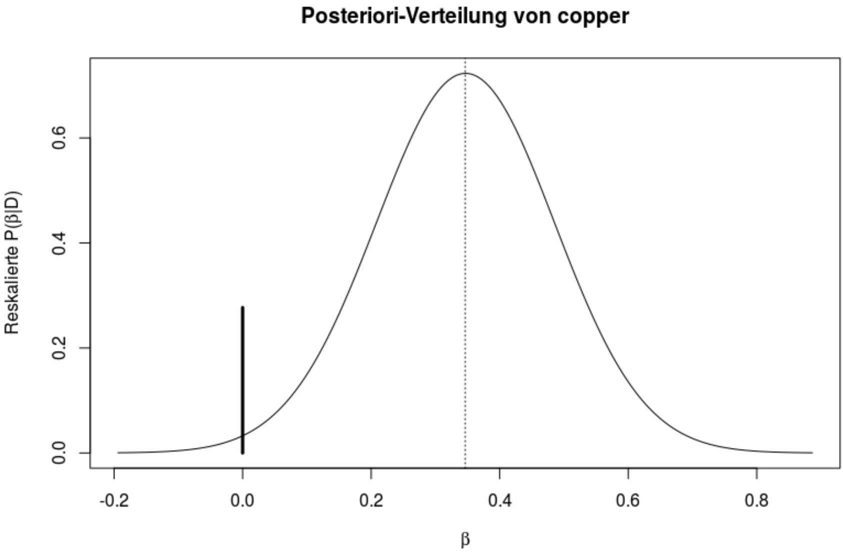
$\Rightarrow$ Kann auch **für** $\beta = 0$ sprechen

Berechnung
Summe der Posteriori-Modellwahrscheinlichkeiten der Modelle, die den Koeffizienten enthalten und nicht auf 0 schätzen

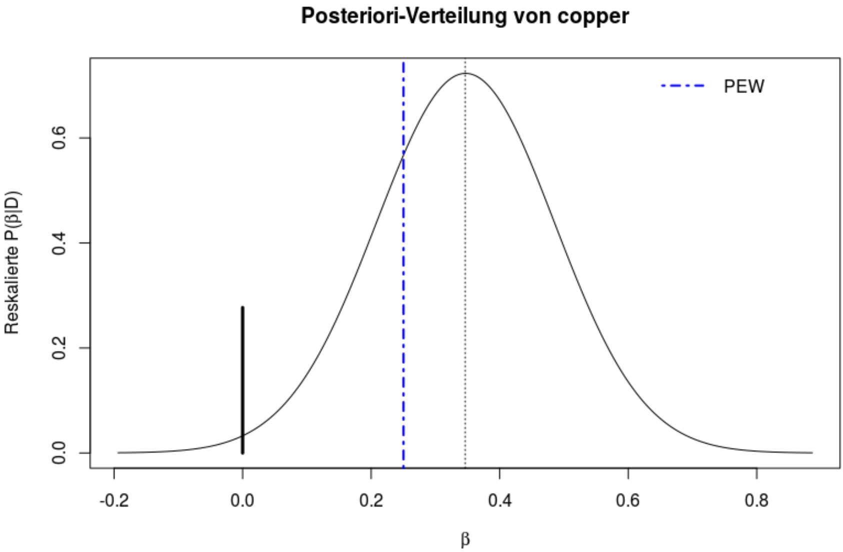
Posteriori-Verteilung von copper



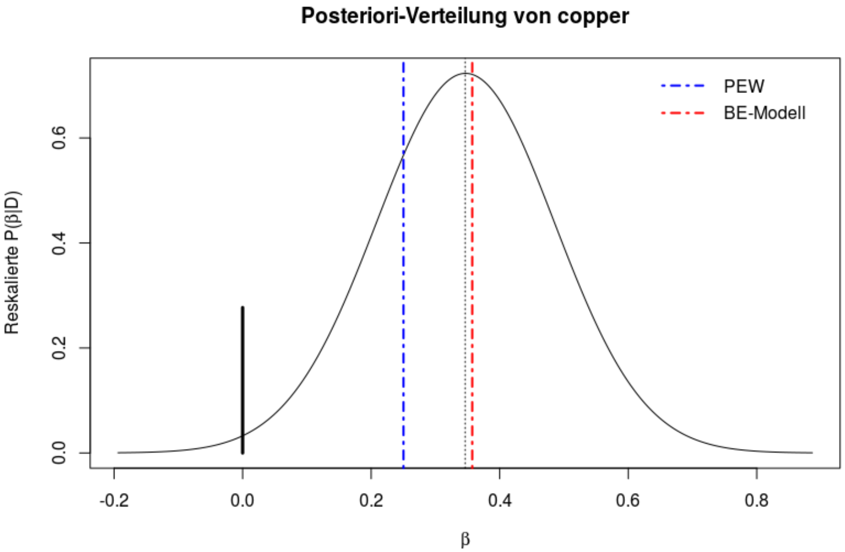
Posteriori-Verteilung von copper



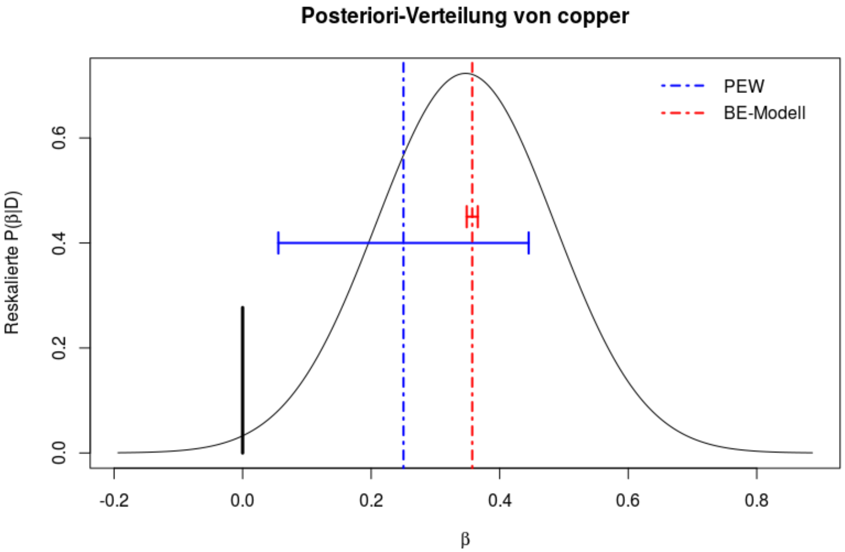
Posteriori-Verteilung von copper



Posteriori-Verteilung von copper



Posteriori-Verteilung von copper



Gliederung

Zusammenfassung

Bayesian Model Averaging

- berücksichtigt mehrere Modelle für die Inferenz
- kann Vorwissen berücksichtigen
- liefert bessere Ergebnisse bei der Prognose

Implementierung im Survival-Kontext

- beschränkt auf Cox-PH-Modelle
- verwendet grobe Approximationen
- berücksichtigt kein Vorwissen

$\Rightarrow$ Beliebtheit durch leichte Grundidee und leichte Implementierung

Referenzen

Fleming, T. R. and Harrington, D. P. (1991). *Counting processes and survival analysis*, John Wiley & Sons, Hoboken.

Held, L. (2008). *Methoden der statistischen Inferenz: Likelihood und Bayes*, Spektrum Akademischer Verlag, Heidelberg.

Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T. (1999). Bayesian model averaging: a tutorial, *Statistical science* **14**(4): 382–401.

Klein, J. P. and Moeschberger, M. L. (2005). *Survival analysis: techniques for censored and truncated data*, Springer, New York.

Raftery, A., Hoeting, J., Volinsky, C., Painter, I. and Yeung, K. Y. (2017). *BMA: Bayesian Model Averaging*. R package version 3.18.7.
URL: *https://CRAN.R-project.org/package=BMA*