

TESTY DOBRÉ SHODY PRO MODEL ZRYCHLENÉHO ČASU V ANALÝZE PŘEŽITÍ

Petr Novák

Klíčová slova: Analýza přežití, testy dobré shody, model zrychleného času

Abstrakt: V příspěvku studujeme regresní modely pro analýzu přežití, věnujeme se především možnostem, jak sestavit testy dobré shody pro model zrychleného času. Porovnáváme je s testy pro Coxův model proporcionálního rizika založenými na teorii čítacích procesů. Na simulovaných datech zkoumáme empirické vlastnosti testů těchto modelů, pozorujeme jejich sílu v závislosti na velikosti sledovaného výběru, typu regresorů a tvaru základního rizika. Hledáme, v jakých situacích je možné dobře rozlišit, podle kterého modelu se data chovají a naopak kdy je rozlišení mezi modely obtížnější.

Keywords: Survival analysis, Goodness-of-fit tests, Accelerated Failure Time Model

Abstract: In present work we study regression models in survival analysis, we focus mainly on options how to perform goodness-of-fit tests for the Accelerated Failure Time model. We compare those methods with existing tests for the Cox proportional hazards model which are based on counting process theory. On simulated data, we study empirical properties of these tests. We compare their empirical power for various sample sizes, covariate types and basic hazard. We try to find cases when it is possible to distinguish between the models well and when not.

1 Regrese v analýze spolehlivosti

Studujeme data reprezentující dobu od začátku pozorování do dosažení nějaké předem definované události - poruchy - v závislosti na vysvětlujících proměnných. Počítáme s nezávislým cenzorováním zprava, tj. že u některých jedinců je pozorování ukončeno před dosažením poruchy. Označíme T_i^* skutečné časy událostí a C_i časy cenzorování. Data máme ve tvaru $(T_i, \Delta_i, \boldsymbol{X}_i)_{i=1}^n$, kde $T_i = min(T_i^*, C_i)$, $\Delta_i = I(T_i \leq C_i)$ a $\boldsymbol{X}_i$ je vektor regresorů.

Dále označme $\alpha_i(t) = lim_{h \to 0} P(t \leq T_i^* < t + h | T_i^* \geq t)/h$ rizikovou funkci. Data se reprezentují také jako čítací procesy, označme $N_i(t) = I(T_i \leq t, \Delta_i = 1)$, $Y_i(t) = I(t \leq T_i)$, intenzity $\lambda_i(t) = Y_i(t)\alpha_i(t)$ a kumulované intenzity $\Lambda_i(t) = \int_0^t \lambda_i(s) ds$. Bylo dokázáno, že $M_i(t) := N_i(t) - \Lambda_i(t)$ jsou za platnosti daného modelu martingaly vzhledem k filtraci

$$F_{t^-} = \sigma \left\{ N_i(s), Y_i(s), X_i, 0 \leq s < t, i = 1, ..., n \right\}$$

(Flemming&Harrington, 1991). Pomocí čítacích procesů se dá přepsat logaritmická věrohodnostní funkce dat a jejím derivováním dle případných parametrů získáváme skórový proces $U(t,\beta)$, pro odhady používáme tento proces až do nějakého času τ, vyššího než je čas poslední události (píšeme $U(\beta) = U(\tau,\beta)$).

2 Nejpoužívanější modely

Srovnáme zde dva ze základních regresních modelů analýzy přežití a možnosti jak provést přílušné testy dobré shody. Nejčastěji používaným je Coxův model proporcionálního rizika (Cox, 1972):

$$\alpha_i(t) = \exp(\boldsymbol{X}_i^T\boldsymbol{\beta})\alpha_0(t), \qquad i = 1, ..., n, \qquad t = [0, \tau],$$

kde $\alpha_0(t)$ je rizikovou funkcí tzv. základního rozdělení. Dalším obvyklým je model zrychleného času (Accelerated Failure Time - AFT, Buckley&James, 1979):

$$\log(T_i^*) = -\boldsymbol{X}_i^T\boldsymbol{\beta} + \epsilon_i, \qquad i = 1, ..., n.$$

kde ϵ_i jsou (*iid*). Pozor, neznáme skutečné hodnoty T_i^*, ale pouze pozorované T_i. Platí $\alpha_i(t) = \alpha_0(e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}t)e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}$, kde $\alpha_0(t)$ je rizikovou funkcí pro veličiny $\exp(\epsilon_i)$. Pro $\alpha_0(t)$ odpovídající Weibullovu rozdělení se modely shodují pro $\boldsymbol{\beta}_C = \delta\boldsymbol{\beta}_A$, kde δ je parametr tvaru Weibullova rozdělení. Oba modely se od sebe odlišují interpretací parametrů, i tím, jak jsou motivovány. V Coxově modelu působí hodnoty kovariát přímo na rizikovou funkci, v AFT modelu regesory způsobují, že virtuálně běží čas pro daný subjekt rychleji nebo pomaleji. Je proto dobré umět rozlišit, podle kterého modelu se data chovají.

Testy dobré shody pro AFT model

Dosazením odhadů $\hat{\boldsymbol{\beta}}$ do rovnice modelu získáme rezidua

$$r_i := \log(T_i) + \boldsymbol{X}_i^T\hat{\boldsymbol{\beta}}.$$

Ta narozdíl od ϵ_i nejsou ani nezávislá ani stejně rozdělená, protože odhady $\hat{\boldsymbol{\beta}}$ jsou založené na celém datovém souboru. Vzhledem k asymptotické konzistenci odhadů $\hat{\boldsymbol{\beta}}$ (Lin et al, 1998) mají ale r_i mít přibližně stejnou střední hodnotu. Pokud máme cenzorovaná data, odhadneme rezidua jako

$$\hat{r}_i := \Delta_i r_i + (1 - \Delta_i)E(\epsilon|\epsilon > r_i^C),$$

kde $E(\epsilon|\epsilon > r_i^C)$ odhadneme jako průměrnou hodnotu všech reziduí necenzorovaných pozorování vyšších než $r_i^C = \log T_i + \boldsymbol{X}_i^T\hat{\boldsymbol{\beta}}$. Rozdělíme data do dvou skupin podle hodnot regresorů a testujeme shodu středních hodnot mezi těmito podvýběry. Použijeme t-test a Wilcoxonův test, kvůli nestejnému rozdělení reziduí budou výsledky pouze přibližné. Vyhodnotíme zde proto empirickou sílu testů v závislosti na velikosti výběru, abychom mohli stanovit, jaká je rychlost asymptotické konvergence.

Testy dobré shody pro Coxův model

Za platnosti Coxova modelu je možné pomocí martingalové dekompozice a centrální limitní věty simulovat proces, který je asymptoticky ekvivalentní skórovému procesu $\tilde{U}(t, \hat{\boldsymbol{\beta}}) = \sum_{i=1}^{n} \boldsymbol{X}_i \hat{M}_i(t)$ (blíže viz Bagdonavičius & Nikulin, 2002). Takto získané replikace pak porovnáme s hodnotou spočítanou z dat. Pro testování použijeme supremovou statistiku $sup_{t\in[0,\tau]}\|\tilde{U}(\hat{\boldsymbol{\beta}}, t)\|$. Pokud její hodnota překročí $(1-\alpha)$% hodnot simulovaných statistik, zamítáme hypotézu, že data se chovají podle Coxova modelu. Vždy jsme vyráběli 1000 replikací.

3 Simulační studie

Generovali jsme data z Coxova i z AFT modelu, jako základní rozdělení bylo použito Gamma rozdělení $\Gamma(a = 1/100, p = 5)$ a Lognormální rozdělení $LN(\mu = 5, \sigma^2 = 1)$. Použili jsme data s jedním regresorem, jednak spojitým s hodnotami generovanými z $N(3, 1)$ a jednak faktorovým s hodnotami 0 a 1 z $Alt(1/2)$. Hodnoty parametru jsme uvažovali $\beta = 1$ a 2 abychom porovnali vliv síly závislosti. Vždy byly zkoumány dvě varianty, bez cenzorování a s nezávislým náhodným cenzorováním (okolo jedné čtvrtiny dat). Byly použity vzorky velikosti 20, 50, 100, 200, 500 a 1000. Na data simulovaná podle Coxova modelu jsme zkoušeli testy AFT modelu a naopak. Zvolili jsme hladinu $\alpha = 0.05$, vždy jsme nagenerovali 1000 opakování a počítali, kolikrát test na této hladině hypotézu zamítne. Tak získáme empirickou sílu proti dané alternativě. Výsledky viz tabulky 1 a 2.

Výsledky - testy Coxova modelu na datech z AFT

- Empirická síla roste s velikostí výběru vždy vyjma případu Gamma rozdělení s faktorovým regresorem a $\beta = 2$.
- Síla vyšší v případech bez cenzorování.
- U lognormálního základního rozdělení je síla vyšší u $\beta = 2$ než u $\beta = 1$, u Gamma rozdělení naopak.
- Při lognormálním rozdělení síla výrazně vyšší při stejném n než při Gamma.

Výsledky - testy AFT na datech z Coxova modelu

- Empirická síla roste s velikostí výběru ve všech případech.
- Síla vyšší v případech bez cenzorování při spojitém regresoru, při faktorovém naopak vyšší s cenzorováním.
- Síla vyšší u $\beta = 2$ než u $\beta = 1$.
- Při lognormálním rozdělení síla výrazně vyšší při stejném n než u Gamma. Použitelný počet zamítnutých výběrů je dosažen u Lognormálního rozdělení pro 200 až 500 pozorování, u Gamma pro 500 až 1000.

Zákl.rozd.	Gamma				Lognormální			
β	1		2		1		2	
Cenzorování	NC	C	NC	C	NC	C	NC	C
Regresor	Spojitý							
20	0.054	0.053	0.071	0.04	0.133	0.103	0.148	0.047
50	0.107	0.09	0.094	0.079	0.291	0.206	0.288	0.204
100	0.234	0.131	0.146	0.112	0.499	0.425	0.423	0.305
200	0.336	0.257	0.249	0.169	0.785	0.614	0.773	0.555
500	0.63	0.552	0.347	0.24	0.995	0.968	0.97	0.84
1000	0.928	0.769	0.557	0.293	1.000	0.997	1.000	0.987
Regresor	Faktorový							
20	0.194	0.26	*0.921*	*0.931*	0.128	0.132	0.225	0.276
50	0.131	0.115	*0.659*	*0.754*	0.166	0.166	0.362	0.300
100	0.245	0.221	0.433	0.532	0.330	0.272	0.562	0.513
200	0.424	0.395	0.203	0.243	0.570	0.508	0.845	0.796
500	0.784	0.696	0.37	0.250	0.904	0.888	0.996	0.990
1000	0.960	0.92	0.661	0.520	0.992	0.984	1.000	1.000

Tabulka 1: Podíl výběrů kde byl Coxův model zamítnut na hladině 0.05 - data z AFT modelu

- Wilcoxonův a t-test srovnatelné u necenzorovaných dat, u cenzorovaných je lepší t-test.
- Celkově nižší síla než u testů Coxova modelu

4 Shrnutí

Aby bylo možné rozlišit, podle kterého z modelů se data chovají, je potřeba v některých případech velký počet pozorování. Testy Coxova modelu vykazují vyšší empirickou sílu než testy pro model zrychleného času. To můžeme přisoudit tomu, že použité metody jsou zde pouze přibližné. Zlepšení by mohlo přinést vyvynutí testů založených na martingalových reziduálech, podobně jako pro Coxův model. Dalším předmětem zkoumání by mohly být i situace s regresory s proměnlivými hodnotami v čase.

Poděkování:
Tato práce byla podporována granty
GAAV No. IAA101120604 a SVV 261315/2010.

Adresa:
MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8,
ÚTIA AV ČR, Pod Vodárenskou věží 4, 182 08 Praha 8

E-mail: `novakp@karlin.mff.cuni.cz`

Zákl.rozd.		Gamma				Lognormální			
β		1		2		1		2	
Cenzorování		NC	C	NC	C	NC	C	NC	C
Regresor		Spojitý							
20	T	0.012	0.011	0.003	0.007	0.052	0.008	0.001	0.007
	W	0.010	0.008	0.003	0.009	0.052	0.007	0	0.003
50	T	0.012	0.008	0.011	0.026	0.014	0.024	0.014	0.019
	W	0.004	0.004	0.004	0.023	0.009	0.016	0.008	0.018
100	T	0.022	0.024	0.022	0.008	0.065	0.034	0.012	0.041
	W	0.017	0.020	0.014	0.016	0.063	0.019	0.011	0.026
200	T	0.071	0.047	0.063	0.035	0.181	0.092	0.147	0.039
	W	0.047	0.027	0.049	0.025	0.163	0.031	0.186	0.028
500	T	0.086	0.058	0.131	0.055	0.653	0.364	0.829	0.390
	W	0.065	0.023	0.151	0.042	0.632	0.205	0.941	0.276
1000	T	0.294	0.182	0.353	0.237	0.890	0.668	0.988	0.666
	W	0.238	0.114	0.356	0.208	0.888	0.402	0.997	0.504
Regresor		Faktorový							
20	T	0	0.003	0.001	0.010	0	0.017	0.002	0.016
	W	0	0	0	0.007	0	0	0	0.004
50	T	0.002	0.004	0.007	0.019	0	0.012	0.02	0.040
	W	0	0.003	0	0.012	0	0.003	0	0.012
100	T	0	0.010	0.021	0.061	0.005	0.027	0.119	0.224
	W	0	0.005	0.003	0.057	0.005	0.008	0.092	0.066
200	T	0.012	0.023	0.109	0.218	0.050	0.106	0.640	0.676
	W	0.002	0.002	0.051	0.215	0.071	0.022	0.619	0.342
500	T	0.076	0.177	0.663	0.849	0.516	0.608	1	0.994
	W	0.047	0.089	0.542	0.875	0.572	0.296	1	0.960
1000	T	0.361	0.580	0.981	0.998	0.966	0.980	1	1
	W	0.224	0.462	0.971	1	0.965	0.801	1	1

Tabulka 2: Podíl výběrů kde byl AFT model zamítnut na hladině 0.05 - data z Coxova modelu. T - t-test, W - Wilcoxonův test

Reference

[1] Buckley J., James I.R.: *Linear regression with censored data*, Biometrika 66, 429–436, 1979.

[2] Cox D.R.: *Regression models and life tables*, J. Roy. Statist. Soc. Ser. B 34, 187–220, 1972.

[3] Fleming T. R., Harrington D. P.: *Counting Processes and Survival Analysis*, Wiley, New York, 1991.

[4] Lin D.Y., Wei L.J., Ying Z.: *Accelerated failure time models for counting processes*, Biometrika 85, 605–618, 1998.

[5] Nikulin M., Bagdonavičius V.: *Accelerated Life Models*, Chapman&Hall, 2002.