

INFORMAČNÍ BULLETIN



České statistické společnosti

Ročník 28, číslo 1, březen 2017

MODEL KONKURENČNÍCH RIZIK S NETRADIČNÍ APLIKACÍ

COMPETING RISKS MODEL WITH A NON-TRADITIONAL APPLICATION

Petr Volf

Adresa: UTIA AV ČR, Praha

E-mail: volf@utia.cas.cz

Abstrakt: Model konkurenčních rizik je aplikován na analýzu časů prvních gólů ve fotbalových zápasech. Konkuruje si latentní časy vstřelení gólu obou týmů. Předpokládá se jejich exponenciální rozdělení, jejich vzájemná závislost je popsána pomocí vhodné kopuly. V příspěvku jsou zpracována data ze dvou ročníků české první fotbalové ligy, 2014–2016.

Klíčová slova: konkurenční rizika, analýza přežití, sportovní statistika.

Abstract: The competing risks scheme is used to the analysis of time to the first goal in a football (soccer) match. The competing random variables are two latent times (of two teams) to scoring. It is assumed that these times are exponentially distributed, their mutual dependence is described by a copula ensuring the model identifiability. As a real example the data from two seasons 2014–2016 of the Czech First League are analyzed and compared.

Keywords: competing risks, survival analysis, sports statistics.

1. Introduction and motivation

The main theme of the present study has arisen from the question how important is the first goal in the football match and, consequently, how strong is the effort of both teams to score first. The problem of the first goal impact has already been addressed in several papers, either directly, as in Nevo and Ritov (2013), or implicitly, in dynamic models of scoring process, as for instance in Dixon and Robinson (1998) and Volf (2009). The latter also contains a brief simulation study showing the dependence of final result on the first goal author and time, for teams with different characteristics. A common feature of such models is the dependence of scoring intensities not only on the characteristics of teams, but also on the match development. As expected, the results of such studies indicate that the reaction to the goal, either scored or obtained, can differ and depends on the teams strengths; not just on factual, but also on psychological strength.

A basic probabilistic model for the final score of a football match, proposed for instance in Maher (1982), consists of two conditionally independent Poisson random variables. More flexible models are obtained for instance by using inflated Poisson models, which are actually the mixtures of Poisson distribution with a discrete distribution giving higher weights to certain more frequent results (as 0:0, 1:0, 1:1). Another generalization can consist in considering changes or/and a time development of model parameters as well as covariates during the match (again for instance Dixon and Robinson, 1998, or Volf, 2009). Yet another way of the basic model improvement consists in considering an explicit form of dependence between both teams scoring distributions. Let us mention here at least two relevant papers. Karlis and Ntzoufras (2003) have employed a special case of the bivariate Poisson distribution. In the same context, McHale and Scarf (2011) have described the dependence with the aid of a copula model. Interesting is the comparison of conclusions of both approaches. While the correlation in the former model is non-negative (by definition), the latter paper concludes that the correlation is negative and is absolutely larger in more competitive matches. It has to be also said that the use of copula in discrete distribution models is not easy technically (and then computationally), because marginal distribution functions are as a rule expressed by sums of point probabilities, not having a reasonably closed form.

In the present contribution we concentrate to the analysis of the distribution of time to the first goal. Consequently, we deal with continuous-type distributions of random times, their dependence is described with the aid of the competing risks scheme. On one hand the use of a copula for two-dimensional continuous distribution can lead to a nicely closed form of the model, on the other hand it is well known that in the competing risks setting the model may not be identifiable. A proof and an example of this phenomenon are given in Tsiatis (1975), some instances of identifiable (or not) models are presented in Basu and Ghosh (1978) – in these classical studies the notion of copula has not been used yet. Therefore we are facing the problem of reasonable copula selection. Fortunately, it has been revealed (cf. Zheng and Klein, 1995) that when the main objective is to estimate the correlation of competing random variables, the selection of copula type is not of crucial importance.

The next section recalls briefly the scheme of competing risks and the problem of possible non-identifiability. The Barnett copula model is formulated and the way of its maximum likelihood evaluation is derived. The main body of the paper is devoted to the application to the analysis of real data from one season of the Czech First League. The results are discussed and

compared with the data from another season, the impact of the first goal on the final match result is examined, too.

2. Competing risks model

The competing risks model assumes that certain events (e.g. a failure of a device) can be caused by K different reasons. Such a situation is then modelled by K (possibly dependent) random variables (random times, as a rule) T_j , $j = 1, \dots, K$, sometimes accompanied by a variable C of random right censoring. C is then independent of all T_j . Let $\bar{F}_K(t_1, \dots, t_K) = P(T_1 > t_1, \dots, T_K > t_K)$ be the joint survival function of $\{T_j\}$. However, instead of the net times T_j we standardly observe just crude data (sometimes called also the identified minimum) $Z = \min(T_1, \dots, T_K, C)$ and the indicator $\delta = j$ if $Z = T_j$, $\delta = 0$ if $Z = C$.

2.1. Problem of identifiability

The data structure described above allows a direct estimation of the distribution of $Z = \min(T_1, \dots, T_K)$, for instance its survival function $S(t) = P(Z > t) = \bar{F}_K(t, \dots, t)$. Further, we can estimate so called incidence densities

$$f_j^*(t) = dP(Z = t, \delta = j) = - \frac{\partial \bar{F}_K(t_1, \dots, t_K)}{\partial t_j} \Big|_{t_1 = \dots = t_K = t}$$

and also their integrals, the cumulative incidence functions, often called also the crude distribution functions. They are estimable consistently by standard survival analysis methods, see for instance Lin (1997). However, in general, from data (Z_i, δ_i) , $i = 1, \dots, N$, it is not possible to identify either marginal or joint distribution of $\{T_j\}$. Tsiatis (1975) has shown that for an arbitrary joint model we can find a model with independent components having the same incidences, i.e. we cannot distinguish between the models. Namely, this independent model is given by cause-specific hazard functions $h_j^*(t) = f_j^*(t)/S(t)$. As a consequence of the Tsiatis (1975) result, it is necessary to make certain functional form assumptions about the type of both marginal and joint distribution in order to identify them. Several such cases are studied in Basu and Ghosh (1978) and in some other papers. More recent results on identifiability can be found for example in Schwarz et al. (2011) dealing with non-parametric setting, or in Escarela and Carriere (2003) considering Frank copula and parametric models.

2.2. Copula models for dependence

In the sequel we shall consider just 2 competing events, i.e. random variables S, T and data $Z_i = \min(S_i, T_i, C_i)$, $\delta_i = 1, 2, 0$, $i = 1, \dots, N$. The notion of copula offers a way how to model multivariate distributions, we prefer here to use it for modelling the joint survival function $\overline{F}_2(s, t)$ of S, T :

$$\overline{F}_2(s, t) = P(S > s, T > t) = C(\overline{F}_S(s), \overline{F}_T(t), \theta), \quad (1)$$

$\overline{F}_S, \overline{F}_T$ are marginal survival functions of S and T , $C(u, v, \theta)$ is a copula, i.e. a two-dimensional distribution function on $[0, 1]^2$, with uniformly distributed marginals U, V . θ is a copula parameter, which is, as a rule, uniquely connected with the correlation of U and V , hence also with the correlation of S and T . It is seen that the use of a copula allows us to model the dependence structure separately from the analysis of marginal distributions. From this point of view, the identifiability of the copula (and its parameter) and of marginal distributions can be considered as two separate steps. Zheng and Klein (1995) proved that when the copula is known, the marginal distributions are estimable consistently (and then the joint distribution, too, from (1)), even in a non-parametric (so quite general) setting. However, in general, the value of θ has to be known. The problem of proper copula choice is analyzed in a set of papers. Let us mention here Kaishev et al. (2007) comparing performances of several copula types. As already mentioned, a common agreement is that the knowledge (or a good estimate) of parameter θ is much more crucial for a reasonable estimation of a joint distribution. As a consequence, because the knowledge of copula type is still an unrealistic supposition, we can try to use a sufficiently flexible class of copulas, as an approximation, and concentrate on reliable estimation of parameters.

2.3. Barnett copula

Tsiatis (1975) in his example considers two competing random variables S and T with exponential distributions and the following marginal and joint survival functions,

$$\overline{F}_S(s) = e^{-\lambda s}, \quad \overline{F}_T(t) = e^{-\mu t}, \quad \overline{F}_2(s, t) = e^{-\lambda s - \mu t - \gamma st}.$$

The example actually uses so called Barnett copula

$$C(u, v) = u \cdot v \cdot \exp(-\theta \cdot \ln u \cdot \ln v), \quad (2)$$

where $\theta \geq 0$. It follows that when U, V are uniform on $[0, 1]$ random variables tied by copula (2), their correlation $\rho(U, V) \leq 0$; $\theta = 0$ means independence

of U and V . Further, when (2) is used for connection of S, T from above, the parameters are related in the following way: $\gamma = \theta \cdot \lambda \cdot \mu$. Figure 1 shows the dependence of correlation on parameters.

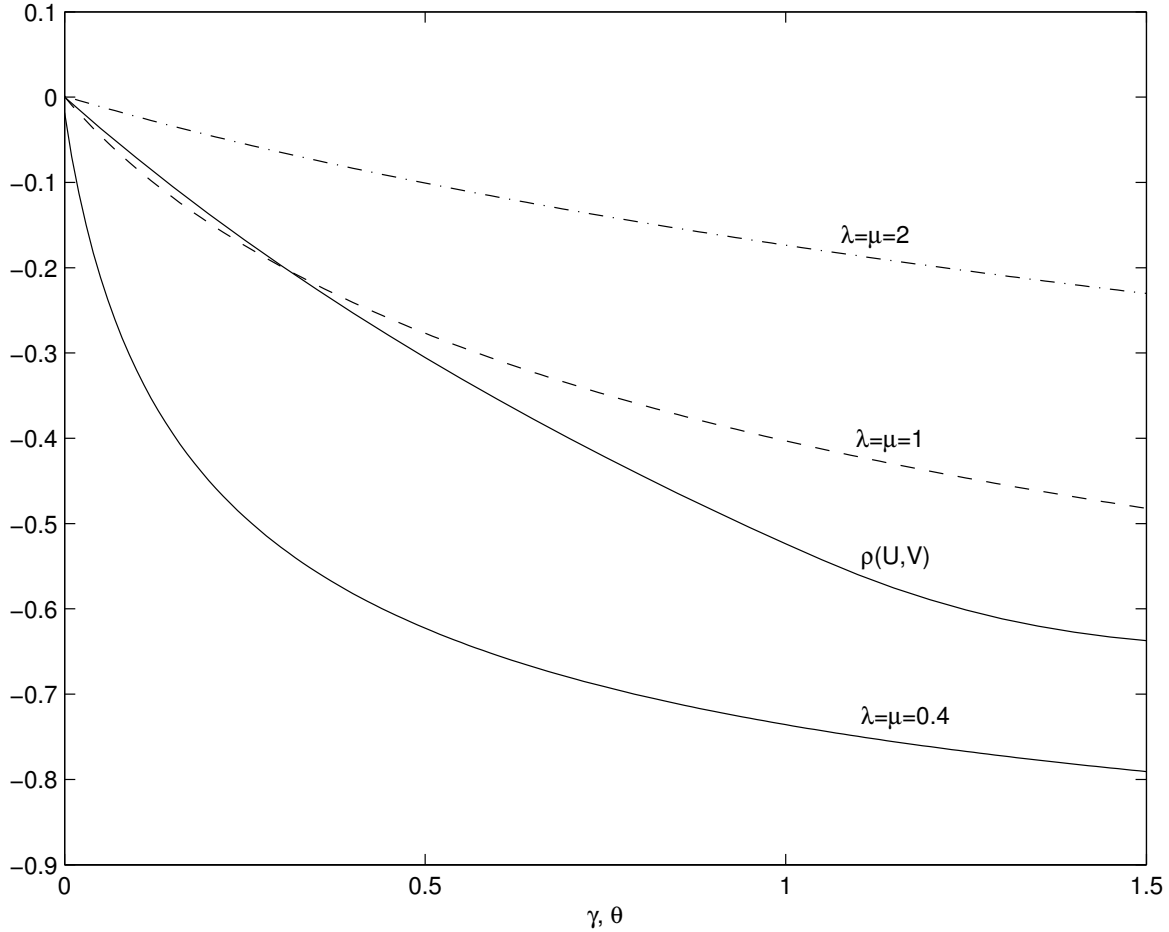


Figure 1: Dependence of $\rho(U, V)$ on parameter θ (solid curve) and $\rho(S, T)$ on γ , for given μ and λ , when $S \sim \exp(\lambda)$, $T \sim \exp(\mu)$.

Identifiability of the model consisting of the Barnett copula (2) and exponential marginal distributions has already been proved by Basu and Ghosh (1978). The model leads to the following likelihood function, given the data $\{z_i, \delta_i, i = 1, \dots, N\}$:

$$\begin{aligned}
 L &= \prod_{i=1}^N \left(\frac{-\partial \bar{F}_2(s, t)}{\partial s} \right)^{[\delta_i=1]} \cdot \left(\frac{-\partial \bar{F}_2(s, t)}{\partial t} \right)^{[\delta_i=2]} \cdot \bar{F}_2(s, t)^{[\delta_i=0]} \Big|_{s=t=z_i} = \\
 &= \prod_{i=1}^N (\lambda + \gamma z_i)^{[\delta_i=1]} \cdot (\mu + \gamma z_i)^{[\delta_i=2]} \cdot \bar{F}_2(z_i, z_i). \quad (3)
 \end{aligned}$$

3. Modelling the time to first goal

We shall use the data from the Czech First League, season 2014–15. Sixteen participating teams played together 240 matches (i.e. twice with each other, home and away). 21 matches ended without goals. More information on the Czech football league can be found on <http://www.sport.cz/fotbal/synot-liga/#vysledky>. Some statistics is provided also in Volf (2016).

As in the standard model of Maher (1982), each team (i) is characterized by its attack parameter a_i and defence parameter b_i , parameter h denotes the advantage of home field. The scoring intensities to the first goal in a match between home team i and away team j are then given as $a_i \cdot b_j \cdot h$, $a_j \cdot b_i$, respectively. Consequently, the time to the 1-st goal arises from two competing exponential random variables

$$S_{ij} \sim \exp(a_i \cdot b_j \cdot h), \quad T_{ij} \sim \exp(a_j \cdot b_i).$$

However, only the incidence of the first of them is observed. Or, in the case of no-score draw (0:0), times are censored by a fixed value, $C_i \equiv 90$ minutes. Further, it was assumed that their mutual dependence can be expressed via the Barnett copula described in Part 2.3.

3.1. Results of analysis

We solved the problem of the maximum likelihood estimation (MLE) of 34 parameters: a_i, b_i of 16 teams, home advantage parameter h , and γ characterizing the dependence. It was assumed that both h and γ were the same in all matches. The results of the MLE of teams parameters are displayed in Table 1. For computational convenience, we estimated $\alpha_i = \ln a_i$, $\beta_i = \ln b_i$, also $\delta = \ln h$. The ML estimates of two common parameters (with half-widths of 95% confidence intervals) were

$$\hat{\delta} = 0.6417 (0.2046), \quad \hat{h} = \exp(\hat{\delta}) = 1.8997, \quad \hat{\gamma} = 0.945 (0.078).$$

The correlation in each particular match depends on the teams parameters and on two common parameters h and γ . Its value can be traced roughly from Figure 1, or computed from the corresponding two-dimensional exponential model. For instance in matches of two leaders, Plzen and Sparta, numerical computation yielded $\rho(S, T) = -0.569$. In a match of teams with rather poor attack and yet fair defence, as for instance for Bohemians and Jihlava, we obtained $\rho(S, T) = -0.800$ which could be interpreted that the first goal was even more important. Further, the value of the parameter $h = 1.9$ indicated that the chance that the home team scores first was about

Team	α		β		a	b
Plzen	0.9742	(0.4664)	-1.8874	(1.7569)	2.6490	0.1515
Sparta	0.3662	(0.6055)	-0.9755	(0.8235)	1.4422	0.3770
Jablonec	0.2115	(0.5584)	-1.4791	(1.1286)	1.2356	0.2278
Ml. Boleslav	0.8080	(0.5539)	-0.2759	(0.6479)	2.2433	0.7589
Pribram	-0.0464	(0.6898)	-0.7362	(0.7491)	0.9547	0.4789
Dukla	-0.2479	(0.8046)	-0.0606	(0.5797)	0.7804	0.9412
Teplice	0.0205	(0.6216)	-1.5465	(1.2794)	1.0207	0.2130
Bohemians	-1.3719	(1.4189)	-0.6103	(0.6467)	0.2536	0.5432
Slovacko	0.2151	(0.6469)	-0.2541	(0.6111)	1.2400	0.7756
Jihlava	-0.2056	(0.7296)	-0.8168	(0.7615)	0.8141	0.4419
Slavia	0.3320	(0.5780)	-0.7249	(0.7983)	1.3938	0.4843
Liberec	-0.5043	(0.8504)	-0.3311	(0.6083)	0.6039	0.7181
Ostrava	-0.3343	(0.7779)	-0.4247	(0.6289)	0.7159	0.6540
Brno	-0.6091	(0.9150)	0.0883	(0.5231)	0.5438	1.0923
Hradec Kr.	-0.3694	(0.8226)	-0.1606	(0.5757)	0.6912	0.8517
C. Budejovice	-0.0435	(0.9000)	0.3128	(0.4921)	0.9574	1.3672

Table 1: **Results:** Estimated parameters $\alpha_i = \ln a_i$, $\beta_i = \ln b_i$ (with half-widths of approximate 95% conf. intervals in brackets), then a_i , b_i . Values of parameters are related to time 90 minutes, in order to keep them in a reasonable scale (and to avoid numerical problems, too).

$1.9/2.9 = 0.66$, while in reality from 219 first goals, 129 were scored by home teams, $129/219 = 0.59$. Let us remark here that the estimated teams parameters are, to certain extent, just relative, that parameter values $a_i \cdot c$, b_i/c yield the same model, for any $c > 0$.

3.2. Brief statistics of the first goal impact

Let us provide also a brief statistics concerning the relationship between the first goal and the final result. From 240 matches of the season 21 ended without goals. From the remaining 219 matches there were 37 draws (other than 0:0), 116 home wins and 66 away wins. Further, from these 219 matches with goals, in 156 cases the team scoring first was also the winner, in 26 cases the opposite had occurred (and 37 ended by draw). Estimated pro-

portion $\hat{p} = 156/219 = 0.7123$ is significantly larger than 0.5 and an approximate (asymptotic) 95% confidence interval for this proportion equals (0.6524, 0.7723). On the other hand, the chance to turn over the score after obtaining the first goal can be described by the estimated proportion $\hat{q} = 26/219 = 0.1187$, which is still significantly larger than zero, yielding a 95% confidence interval (0.0759, 0.1616).

	Order 2014–15	a(14)	b(14)	a(15)	b(15)	Order 2015–16
1	Plzen	2.6490	0.1515	1.8218	0.2592	1
2	Sparta	1.4422	0.3770	1.0447	0.6478	2
3	Jablonec	1.2356	0.2278	0.7470	0.5481	7
4	Ml. Boleslav	2.2433	0.7589	1.5209	0.6408	4
5	Pribram	0.9547	0.4789	0.6197	0.9066	14
6	Dukla	0.7804	0.9412	1.2129	0.5980	10
7	Teplice	1.0207	0.2130	0.8845	0.8982	12
8	Bohemians	0.2536	0.5432	0.8931	0.5389	9
9	Slovacko	1.2400	0.7756	0.3463	1.0957	8
10	Jihlava	0.8141	0.4419	0.7611	0.5699	11
11	Slavia	1.3938	0.4843	1.4646	0.3974	5
12	Liberec	0.6039	0.7181	1.0913	0.5251	3
13	Ostrava	0.7159	0.6540	1.3502	1.1322	16
14	Brno	0.5438	1.0923	0.6325	0.5334	6
15	H. Kr./Olomouc	0.6912	0.8517	0.4542	0.4836	15
16	C. Budej./Zlin	0.9574	1.3672	1.1406	0.9276	13

Table 2: Comparison of results, i.e. estimated parameters and final order of teams in seasons 2014–15 and 2015–16.

3.3. Comparison with season 2015–16

The purpose of this section is to examine how stable are the parameters during a longer period. The data from the Czech First League season 2015–16 have been analyzed in the same manner as the season previous. Table 2 displays estimated teams parameters (with two new teams). It is seen that, at least for 5 teams, their final order as well as their parameters differ signifi-

cantly from the values in Table 1. Hence, it has no much sense to use both data-sets jointly, we shall obtain some average values with poor interpretation. As regards two common parameters, we have obtained

$$\hat{\delta} = 0.4837 (0.1966), \quad \hat{h} = \exp(\hat{\delta}) = 1.6221, \quad \hat{\gamma} = 1.450 (0.117).$$

They are not far from the former values, nevertheless, from the statistical point of view, estimated γ differs significantly (corresponding confidence intervals are disjoint). We guess that the results reflect a specific feature of the season 2015–16: it was rather rich to scored goals: 676 goals were scored, while in the preceding season their number was 645. Consequently, a smaller correlations could be expected. Regarding the correlation of exponential distributions in the framework of the Barnett copula model, it is influenced jointly by teams parameters, by h , and by γ . Then, though the γ is larger, the other parameters may make the correlations smaller (compare also with figure 1). For instance, now the match of Sparta and Plzen leads to estimated $\rho = -0.602$, for Bohemians against Jihlava the value $\rho = -0.676$ has been obtained. Notice that the first-goal intensity of the Bohemians has increased considerably.

Let us again summarize the first goals statistics. From 240 matches of the 2015–16 season just 12 ended without goals. From remaining 228 matches there were 46 draws (other than 0:0), 119 home and 63 away wins. The home team has scored first in 134 cases, the away team in 94 cases. The proportion characterizing the home advantage equals then $134/228 = 0.59$, while the home advantage parameter leads to $\hat{h}/(1 + \hat{h}) = 1.62/2.62 = 0.62$, which is a comparable value. Further, from 228 matches with goals, in 146 cases the team scoring first was also the winner, in 36 cases the the first scoring team has lost. Estimated proportions are now $\hat{p} = 146/228 = 0.6404$ with asymptotic 95% confidence interval $(0.5785, 0.7026)$, and $\hat{q} = 36/228 = 0.1579$ $(0.1106, 0.2052)$. Both are comparable with proportions from the season 2014–15.

4. Discussion and concluding remarks

The results lead to a conclusion that the correlation is, as a rule, negative, and is absolutely larger in more competitive matches, i.e. the matches of teams with good defence and comparable attack abilities. This conclusion is thus comparable with the results of McHale and Scarf (2011). The analysis has been repeated with the use of Gauss copula (Volf, 2016). Here, the correlation was common for all matches, i.e. not so flexible. Nevertheless, its estimated value was negative, too, and teams parameters comparable with

those reported here. It has to be pointed out that the parameters estimated above concern just the stage of a match up to the first goal. It can be expected that the team performance changes during the match and is related to the actual score, elapsing time, and to other factors characterizing the match state. This aspect is also reflected by more advanced models of score development; see again for instance Dixon and Robinson (1998), Volf (2009) and an overview of models provided there. Hence, the approach proposed in the present study can be extended to the analysis of times to next goals. Another generalization can consider different copula parameters for certain groups of matches or teams.

In spite of the fact that the identification of proposed model, and therefore also the consistency of parameter estimates, are guaranteed theoretically, simulated experiments show rather slow convergence of estimates to ‘true’ values. It is also seen that the confidence intervals for parameters (approximate, i.e. based on asymptotic normality of estimates) are rather wide, which is a natural consequence of a rather high ratio of the number of parameters to the number of matches. Simultaneously, however, this follows from the fact that the log-likelihood function, as a function of γ , is rather flat.

References

- [1] Basu A. P., Ghosh J. K. (1978): Identifiability of the Multinormal and Other Distributions under Competing Risks Model. *Journal of Multivariate Analysis* **8**, 413–429.
- [2] Dixon M. J., Robinson M. E. (1998): A birth process model for association football matches. *The Statistician* **47**, 523–538.
- [3] Escarela G., Carriere J. F. (2003): Fitting competing risks with an assumed copula. *Statistical Methods in Medical Research* **12**, 333–349.
- [4] Kaishev V. K., Dimitrova D. S., and Haberman S. (2007): Modelling the joint distribution of competing risks survival times using copula functions. *Insurance: Mathematics and Economics* **41**, 339–361.
- [5] Karlis D., Ntzoufras I. (2003): Analysis of sports data by using bivariate Poisson models. *J. R. Stat. Soc. Ser. D* **52**, 381–394.
- [6] Lin D. Y. (1997): Non-parametric inference for cumulative incidence functions in competing risks studies. *Statistics in Medicine* **16**, 901–910.
- [7] Maher M. J. (1982): Modelling association football scores. *Stat. Neerl.* **36**, 109–118.
- [8] McHale I., Scarf P. A. (2011): Modelling the dependence of goals scored by opposing teams in international soccer matches. *Statistical Modelling* **11**, 219–236.

- [9] Nevo D., Ritov Y. (2013): Around the goal: examining the effect of the first goal on the second goal in soccer using survival analysis methods. *Journal of Quantitative Analysis in Sports* **9**, 165–177.
- [10] Schwarz M., Jongbloed G., and Van Keilegom I. (2013): On the identifiability of copulas in bivariate competing risks models. *Canadian Journal of Statistics* **41**, 291–303
- [11] Tsiatis A. (1975): A nonidentifiability aspects of the problem of competing risks. *Proc. Nat. Acad. Sci. USA* **72**, 20–22.
- [12] Volf P. (2009): A random point process model for the score in sport matches. *IMA Journal of Management Mathematics* **20**, 121–131.
- [13] Volf P. (2016): On Statistical Analysis of Competing Risks with Application to the Time of First Goal. *Journal of Mathematics and Statistical Science* **2**, 606–623.
- [14] Zheng M., Klein J.P. (1995): Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika* **82**, 127–138.

DIAGNOSTICKÉ METODY PRO MODEL ZRYCHLENÉHO ČASU

DIAGNOSTIC METHODS FOR THE ACCELERATED FAILURE TIME MODEL

Petr Novák

Adresa: FIT ČVUT v Praze, Thákurova 9, 160 00 Praha 6

E-mail: petr.novak@fit.cvut.cz

Abstrakt: Regresní model zrychleného času umožňuje interpretovat závislost doby přežití jedince na vysvětlujících proměnných za přítomnosti cenzorování. V příspěvku zkoumáme diagnostické metody pro tento model a rozšiřujeme možnosti testování dobré shody modelu s daty. Využíváme jak postupy založené na reziduích lineární regrese modifikovaných pro cenzorovaná data, tak teorii čítacích procesů. Na simulovaných příkladech zkoumáme vlastnosti popsaných metod v různých situacích.

Klíčová slova: Analýza přežití, regresní modely, testy dobré shody.

Abstract: The Accelerated failure time model allows us to interpret the dependence of the life time distribution of an individual on available explanatory variables in the presence of censoring. In this work we explore diagnostic methods and present new goodness-of-fit testing procedures for this model. We use methods based on classic linear regression residuals approach modified for censored data as well as counting process theory. We examine the properties of presented methods on simulated examples.

Keywords: Survival analysis, regression models, goodness-of-fit testing.

1. Úvod

Při zkoumání rozdělení doby přežití v medicínských studiích či doby do poruchy v technické praxi a její závislosti na případných vysvětlujících proměnných je klíčové zvolit vhodný regresní model. Používané metody se opírají často o rizikovou funkci a umožňují vzít v potaz i cenzorovaná data, kdy pozorovaní jedinci mohou opustit studii dříve, než dojde k úmrtí či selhání. Vedle nejobvyklejšího Coxova modelu proporcionálních rizik [5] a Aalenova aditivního modelu [2] byl navržen i model zrychleného času (Accelerated Failure Time – AFT, [4]), kde hodnoty regresorů způsobují, že virtuální čas každého jedince může plynout rychleji či pomaleji než čas pozorovaný. V této práci se věnujeme AFT modelu, možnostem jak diagnostikovat odlehlá či

vlivná pozorování a především metodám testů dobré shody. Postupy zakládáme na faktu, že odhady komponent AFT modelu je možné provádět jak pomocí metod klasické lineární regrese, tak pomocí teorie čítacích procesů, přičemž oba postupy dávají asymptoticky shodné odhady.

2. Model zrychleného času

Pozorujeme $i = 1, \dots, n$ jedinců. Označme \mathcal{T}_i skutečné časy selhání a C_i časy cenzorování, které uvažujeme nezávislé na \mathcal{T}_i . Pozorujeme pak časy konce sledování $T_i = \min(\mathcal{T}_i, C_i)$, indikátory necenzorovaných událostí $\Delta_i = I(\mathcal{T}_i \leq C_i)$ a regresory \mathbf{Z}_i . Předpokládáme, že \mathcal{T}_i jsou nezávislé, spojitě, s hustotami $f_i(t)$, distribučními funkcemi $F_i(t)$, funkcemi přežití $S_i(t) = 1 - F_i(t)$, rizikovými funkcemi

$$\alpha_i(t) = \lim_{h \rightarrow 0^+} \frac{P(t \leq \mathcal{T}_i < t + h | \mathcal{T}_i \geq t)}{h}$$

a kumulovanými rizikovými funkcemi $A_i(t) = \int_0^t \alpha_i(s) ds$.

V modelu modelu zrychleného času uvažujeme log-lineární závislost:

$$\log \mathcal{T}_i = -\mathbf{Z}_i^T \boldsymbol{\beta} + \epsilon_i, \quad \epsilon_i \sim \text{i.i.d.}$$

Tento vztah si můžeme představit jako transformaci času mezi pozorovaným a virtuálním časem $t \rightarrow t e^{\mathbf{Z}_i^T \boldsymbol{\beta}}$. Dá se ukázat, že riziková funkce má pak tvar $\alpha_i(t) = e^{\mathbf{Z}_i^T \boldsymbol{\beta}} \alpha_0(t e^{\mathbf{Z}_i^T \boldsymbol{\beta}})$, kde $\alpha_0(t)$ je základní riziková funkce odpovídající veličinám $\exp(\epsilon_i)$. Pro Weibullovo základní rozdělení je AFT shodný s Coxovým modelem.

Uvažujme semiparametrický model, kde základní riziko α_0 je neznámé. Zaměříme se na odhad parametrů $\boldsymbol{\beta}$. Porovnáme konstrukci odhadů pomocí metody maximální věrohodnosti s využitím teorie čítacích procesů, viz [10], s metodou nejmenších čtverců s korekcí pro cenzorovaná pozorování dle [4].

2.1. Metoda I – Maximální věrohodnost a čítací procesy

Data lze přepsat pomocí čítacích procesů na transformované škále času:

$$N_i^*(t, \boldsymbol{\beta}) = \Delta_i I(e^{\mathbf{Z}_i^T \boldsymbol{\beta}} T_i \leq t), \quad Y_i^*(t, \boldsymbol{\beta}) = I(e^{\mathbf{Z}_i^T \boldsymbol{\beta}} T_i \geq t),$$

kde N_i^* je indikátor necenzorované události do času t a Y_i^* indikátor, zda je v čase t i -tý jedinec ještě v riziku. Věrohodnost dat se skládá z částí pro necenzorované a cenzorované události. Použitím $\alpha_i(t) = \frac{f_i(t)}{S_i(t)}$ a $S_i(t) = e^{-A_i(t)}$

lze věrohodnost zapsat ve tvaru:

$$L = \prod_{i=1}^n (f_i(T_i))^{\Delta_i} (S_i(T_i))^{1-\Delta_i} = \prod_{i=1}^n (\alpha_i(T_i))^{\Delta_i} \exp(-A_i(T_i)).$$

Logaritmujeme, dosadíme rizikovou funkci AFT modelu a prepíšeme pomocí čítacích procesů. Derivováním získáme skóre, které ale záleží na neznámé kumulativní základní rizikovou funkci A_0 . Když ji nahradíme Nelson-Aalenovým odhadem $\hat{A}_0(t) = \int_0^t \frac{dN_i^*(s, \beta)}{\sum_i Y_i^*(s, \beta)}$, viz [1], dostaneme přibližné skóre

$$\tilde{U}(\beta, w_0) = \sum_{i=1}^n \int_0^\tau w_0(s) \left(\mathbf{z}_i - \frac{\sum_{k=1}^n Y_k^*(s, \beta) \mathbf{z}_k}{\sum_{k=1}^n Y_k^*(s, \beta)} \right) dN_i^*(s, \beta), \quad (1)$$

kde $w_0(s) = (1 + \frac{s\alpha'(s)}{\alpha(s)})$ je váhová funkce a τ je omezující konstanta vyšší než všechna pozorování (viz [10]). Dá se ukázat, že místo w_0 lze použít pro odhad libovolnou funkci splňující určité předpoklady, např. $w_1(s) \equiv 1$, viz [10]. Protože přibližné skóre není spojitě v β , odhady parametrů pak získáme minimalizací $\|\tilde{U}(\beta, w)\|$. Za určitých podmínek regularity pak platí $\sqrt{n}(\hat{\beta} - \beta_0) \rightarrow \mathcal{N}(\mathbf{0}, \Sigma_w)$, přičemž rozptyl odhadů závisí na použité váhové funkci $w(t)$.

2.2. Metoda II – Lineární regrese s korekcí odezvy

AFT model je modelem cenzorované lineární regrese. Pro jednoduchost zavedme $\mathcal{Y}_i = \log T_i$ pro skutečná data a $Y_i = \log T_i$ pro cenzorovaná. Získáme běžný model $\mathcal{Y}_i = -\mathbf{z}_i^T \beta + \epsilon_i$, ve kterém navíc uvažujeme absolutní člen a nulovou střední hodnotu odchylek ϵ_i . Kdybychom neuvažovali cenzorování, odhadli bychom parametry metodou nejmenších čtverců

$$\hat{\beta} = -(\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{y}.$$

Popíšeme zde postup sestavení odhadů za přítomnosti cenzorování. Místo cenzorovaných pozorování použijeme jejich odhad, viz [4]. Uvažujme rezidua

$$r_i = Y_i + \mathbf{z}_i^T \hat{\beta}.$$

Pro cenzorovaná pozorování neposkytují rezidua aproximaci odchylek ϵ_i , protože jsou systematicky nižší. Použijeme korekci

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

kde $E(\epsilon|\epsilon > r_i)$ odhadneme jako

$$\widehat{E}(\epsilon|\epsilon > r_i) = \frac{\sum_{j:r_j > r_i} \Delta_j r_j d\widehat{F}_0(r_j)}{\widehat{S}_0(r_i)},$$

přičemž \widehat{S}_0 je Kaplan-Meierův odhad funkce přežití odpovídající základnímu rozdělení $\widehat{S}_0(t) = \prod_{j:r_j \leq t} \left(\frac{n - \text{rank}(r_j)}{n - \text{rank}(r_j) + 1} \right)^{\Delta_j}$. Tím lze odhadnout skutečné doby událostí jako

$$\widehat{Y}_i = \begin{cases} Y_i & \text{non-cens.} \\ -\mathbf{Z}_i^T \widehat{\boldsymbol{\beta}} + \widehat{r}_i & \text{cens.} \end{cases}$$

Hledáme $\boldsymbol{\beta}$ pomocí metody nejmenších čtverců s korigovanými odezvami, tedy řešíme rovnici

$$\mathbf{Z}^T (\widehat{\mathbf{Y}} + \mathbf{Z}^T \boldsymbol{\beta}) = \mathbf{0}.$$

Obecněji bychom mohli uvažovat i minimalizaci vážených čtverců

$$\Psi(\boldsymbol{\beta}, s) = \sum_{i=1}^n \mathbf{Z}_i^T s(\widehat{Y}_i + \mathbf{Z}_i^T \boldsymbol{\beta}) = \mathbf{0}, \quad (2)$$

kde $s(t)$ je váhová funkce, výše $s(t) = t$. Korigovaná odezva se dá zapsat vektorově dle [3] jako:

$$\widehat{\mathbf{Y}} = -\mathbf{Z}^T \boldsymbol{\beta} + \mathbf{Q}(\mathbf{Y} + \mathbf{Z}^T \boldsymbol{\beta}),$$

kde \mathbf{Q} je matice s prvky

$$q_{ij} = \begin{cases} \Delta_i & i = j \\ (1 - \Delta_i) \frac{\Delta_j r_j d\widehat{F}_0(r_j)}{1 - \widehat{F}_0(r_i)} & r_i < r_j \\ 0 & \text{jinak.} \end{cases}$$

Dosadíme do řešené rovnice a získáme (z $\mathbf{Q}\mathbf{Q} = \mathbf{Q}$ a $\mathbf{Q}\mathbf{Y} = \mathbf{Q}\mathbf{Y}$, viz [3]):

$$\widehat{\boldsymbol{\beta}} = -(\mathbf{Z}^T \mathbf{Q} \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{Q} \mathbf{Y}.$$

Protože \mathbf{Q} závisí na $\boldsymbol{\beta}$, je nutno postupovat iterativně. V každém kroku nejprve dosadíme $\widehat{\boldsymbol{\beta}}^{(m)}$ do \mathbf{Q} , spočítáme $\widehat{\boldsymbol{\beta}}^{(m+1)}$ a opakujeme. Lze ukázat, že za platnosti modelu bude postup konvergovat, viz [9].

2.3. Souvislost mezi metodami

Odhady metodou maximální věrohodnosti získáme minimalizací normy skóre $\tilde{U}(\boldsymbol{\beta}, w)$ z (1). Oproti tomu odhady metodou vážených nejmenších čtverců z (2) získáme řešením $\Psi(\boldsymbol{\beta}, s) = \mathbf{0}$. Dá se ukázat, viz [8], že za platnosti modelu a určitých podmínek regularity je

$$\frac{1}{\sqrt{n}}\Psi(\boldsymbol{\beta}, s) = \frac{1}{\sqrt{n}}\tilde{U}(\boldsymbol{\beta}, w) + o_P(1), \quad (3)$$

pokud zvolíme pro skóre váhovou funkci

$$w(t) = s(t) - \frac{\int_t^\infty s(u)dF_0(u)}{1 - F_0(t)}.$$

Odhady $\hat{\boldsymbol{\beta}}_{MLE}$ a $\hat{\boldsymbol{\beta}}_{LS}$ mají tedy stejné asymptotické vlastnosti. Ať získáme odhady jednou či druhou metodou, $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ má asymptoticky nulovou střední hodnotu a konečný rozptyl, který lze odhadnout např. pomocí resamplingu skóre z MLE, viz [6]. Váhovou funkci ve skóre odpovídající $s(t) = t$ můžeme bez změny asymptotických vlastností odhadů nahradit odhadem $\hat{w}(t) = t - \frac{\int_t^\infty ud\hat{F}_0(u)}{1 - \hat{F}_0(t)}$.

3. Regresní diagnostika

Souvislost metod odhadů lze použít pro diagnostiku modelu. Nastíníme zde postupy hledání vlivných pozorování a testování dobré shody modelu s daty.

3.1. Detekce vlivných pozorování

Rezidua modelu lze zapsat jako

$$\hat{\mathbf{r}} = \hat{\mathbf{y}} + \mathbf{Z}^T \hat{\boldsymbol{\beta}} = (\mathbf{I} - \mathbf{Z}^T (\mathbf{Z}^T \mathbf{Q} \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{Q}) \hat{\mathbf{y}}.$$

Tedy získáme hat-matici $\mathbf{H} = \mathbf{Z}^T (\mathbf{Z}^T \mathbf{Q} \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{Q}$ a \mathbf{M} matici $\mathbf{M} = \mathbf{I} - \mathbf{H}$, které lze využít k regresní diagnostice. Např. podle diagonálních prvků (leverage)

$$h_{ii} = \mathbf{z}_i^T (\mathbf{Z}^T \mathbf{Q} \mathbf{Z})^{-1} \mathbf{z}_i,$$

kde $q_{\bullet i}$ značí i -tý sloupec matice \mathbf{Q} , lze označit za vlivná pozorování taková, pro která je $h_{ii} > 2(p + 1)/n$, viz [9]. Pro cenzorovaná pozorování je $h_{ii} = 0$, takže jejich případný vliv tímto postupem detekovat nelze. \mathbf{H} závisí na odhadnutých parametrech prostřednictvím matice \mathbf{Q} , tedy výsledky budou jen přibližné. Byly navrženy i obdoby dalších diagnostických metod, jako například Cookova vzdálenost, viz [3].

3.2. Testy dobré shody – vzdálenost odhadů

Za platnosti modelu by obě metody měly dle (3) dávat blízké odhady. Můžeme tedy odhadnout parametry oběma způsoby a testovat, zda nejsou významně vzdálené. Využijeme replikace odhadů pomocí bootstrapového resamplingu skóre dle [6] pro MLE a dle [8] pro LS. Můžeme tak generovat $\hat{\beta}_{MLE}^G$ a $\hat{\beta}_{LS}^G$, pro která má $\sqrt{n}(\hat{\beta}_{MLE}^G - \hat{\beta}_{LS}^G)$ asymptoticky stejné rozdělení jako $\sqrt{n}(\hat{\beta}_{MLE} - \hat{\beta}_{LS})$. Z replikovaných hodnot odhadneme varianční matici rozdílu odhadů $\Sigma_{\hat{\beta}_{MLE} - \hat{\beta}_{LS}}$ a pro testování použijeme statistiku

$$W = (\hat{\beta}_{MLE} - \hat{\beta}_{LS})^T \hat{\Sigma}_{\hat{\beta}_{MLE} - \hat{\beta}_{LS}}^{-1} (\hat{\beta}_{MLE} - \hat{\beta}_{LS}), \quad (4)$$

která bude mít za platnosti modelu χ_p^2 rozdělení.

3.3. Testy dobré shody – rezidua

Zkoumáme, zda je závislost doby do selhání na určité proměnné popsána modelem dobře. Můžeme například rozdělit data do dvou skupin podle mediánu zkoumané proměnné a spočítat korigovaná rezidua \hat{r}_i . Za platnosti modelu by rezidua v obou skupinách měla mít shodnou střední hodnotu. Kdybychom znali skutečnou hodnotu parametrů β a neuvažovali cenzorování, mohli bychom použít vhodný dvouvýběrový test shody středních hodnot, např. t -test. Takový postup byl pro AFT model zkoumán v [7], pro některé případy ale nedával příliš silné výsledky.

Pokusíme se test zefektivnit zohledněním vlivu odhadnutých parametrů a korekce cenzorování. Pro test použijeme statistiku

$$T = \sqrt{n} \left(\frac{1}{n_1} \sum_{s_1} \hat{r}_i - \frac{1}{n_2} \sum_{s_2} \hat{r}_i \right), \quad (5)$$

která bude mít za platnosti modelu nulovou střední hodnotu a bude konvergovat k normálnímu rozdělení. Pomocí bootstrapových replikací odhadů parametrů $\hat{\beta}^G$ dle [6] a $G_1, \dots, G_n \sim N(0, 1)$ (i.i.d.) sestavíme statistiku

$$\begin{aligned} \hat{T} &= \sqrt{n} \left(\frac{1}{n_1} \sum_{s_1} \Delta_i (\hat{Y}_i + \mathbf{Z}_i^T \hat{\beta}) G_i - \frac{1}{n_2} \sum_{s_2} \Delta_i (\hat{Y}_i + \mathbf{Z}_i^T \hat{\beta}) G_i \right) \\ &+ \sqrt{n} \left(\frac{1}{n_1} \sum_{s_1} \Delta_i \mathbf{Z}_i^T - \frac{1}{n_2} \sum_{s_2} \Delta_i \mathbf{Z}_i^T \right) (\hat{\beta} - \hat{\beta}^G). \end{aligned}$$

Dá se ukázat, že \hat{T} má asymptoticky stejné rozdělení jako T . Tudíž ji můžeme replikovat, spočítat rozptyl a rozhodnout tak, zda je rozdíl mezi rezidui významně vysoký.

4. Simulační studie

Na základě generovaných dat ověříme empirické vlastnosti navržených testů dobré shody. Simulovali jsme data z AFT modelu samotného, z Coxova modelu a ze špatně specifikovaného AFT modelu, kde byla náhodně vybraná polovina regresorů zdvojnásobena. Základní rozdělení bylo použito gamma $\Gamma(p=20, \lambda=4/100)$ a lognormální $LN(\mu=5, \sigma^2=1)$, s jedním spojitým regresorem $Z_i \sim N(3, 1)$ a $\beta = 2$. Uvažovali jsme jak necenzorovaná data (NC), tak případ se zhruba čtvrtinou nezávisle cenzorovaných pozorování (C). Generovali jsme vzorky o 50 a 500 pozorováních, pro replikace parametrů bylo použito 50 hodnot. Každý případ jsme zopakovali 500krát a pozorovali jsme podíl zamítnutí AFT modelu na hladině 5 %, abychom získali empirickou hladinu a sílu proti daným alternativám. Použili jsme přibližný t -test shody středních hodnot reziduí a testy pomocí statistik W dle (4) a T dle (5).

Tabulka 1: Podíl výběrů zamítnutých různými testy AFT modelu: t -test aplikovaný na rezidua, W – test porovnáním odhadů MLE a LS dle (4), T – test dle (5). NC/C – necenzorované/cenzorované výběry.

Model	Základní rozdělení	n	t -test		W		T	
			NC	C	NC	C	NC	C
AFT	Gamma	50	0,000	0,014	0,000	0,024	0,000	0,004
		500	0,000	0,000	0,006	0,026	0,000	0,000
	LN	50	0,006	0,000	0,012	0,010	0,012	0,014
		500	0,004	0,004	0,036	0,018	0,062	0,060
Cox	Gamma	50	0,000	0,016	0,020	0,052	0,000	0,022
		500	0,000	0,012	0,046	0,000	0,000	0,015
	LN	50	0,000	0,020	0,064	0,022	0,004	0,016
		500	0,046	0,000	0,556	0,000	0,592	0,026
AFT mis-spec.	Gamma	50	0,198	0,022	0,404	0,412	0,608	0,512
		500	0,924	0,970	1,000	1,000	0,934	0,980
	LN	50	0,108	0,034	0,306	0,168	0,658	0,632
		500	0,720	0,700	1,000	0,972	0,794	0,980

V tabulce 1 vidíme, že první i druhý test nepřekračují empirickou hladinu 5 %, tedy nezamítají AFT model samotný ve více než 5 % případů. Někdy je podíl zamítnutí výrazně nižší, což svědčí o jisté konzervativnosti testů.

Poslední test v některých případech hladinu mírně překračuje, což může být dáno nepřesností asymptotiky. Coxův model je nejčastěji zamítán pro lognormální rozdělení bez cenzorování, v ostatních případech Coxův model testy od AFT modelu nerozpoznají. To může být způsobeno chováním chvostu rozdělení, kde se gamma rozdělení podobá Weibullovu, pro které AFT a Coxův model splývají a naopak lognormální rozdělení se odlišuje. Model se špatně specifikovaným regresorem byl zamítnut ve velké části případů. Nově navržené testy dávají ve většině situací lepší výsledky než jednoduchý t -test aplikovaný na rezidua. S cenzorováním empirická síla až na výjimky klesá.

5. Závěr

Prozkoumali jsme dvě možnosti odhadů parametrů modelu zrychleného času, jednak věrohodnostní přístup využívající teorii čítacích procesů a jednak metodu nejmenších čtverců modifikovanou pro cenzorovaná data. Mezi metodami existuje přímá souvislost a asymptotické vlastnosti odhadů jsou za platnosti modelu shodné. Toho jsme využili pro sestavení testů dobré shody pro zkoumaný model. Vlastnosti testů jsme ověřili na simulovaných datech

Dalším krokem by mohlo být zkoumání širší diagnostické využitelnosti reziduí obdobně jako pro necenzorovanou lineární regresi a studium jejich empirických vlastností, případně rozšíření metod pro regresory proměnné v čase.

References

- [1] Aalen O. O. (1975): *Statistical inference for a family of counting processes*. Ph.D. thesis, Univ. of California, Berkeley, 1975.
- [2] Aalen O. O. (1980): A model for non-parametric regression analysis of counting processes. *Lecture Notes in Statist.* 2, 1–25. Springer, NY, 1980.
- [3] Aziz N. (2010): *Analysis and Diagnostics for Censored Regression and Multivariate data*. Ph.D. thesis, Victoria University of Wellington, 2010.
- [4] Buckley J., James I. R. (1979): Linear regression with censored data. *Biometrika* **66**, 429–436.
- [5] Cox D. R. (1972): Regression models and life tables. *J. Roy. Statist. Soc. Ser. B* **34**, 187–220.
- [6] Lin D. Y., Wei L. J. and Ying Z. (1998): Accelerated failure time models for counting processes. *Biometrika* **85**, 605–618.
- [7] Novák P. (2010): Testy dobré shody pro model zrychleného času v analýze přežití. *Informační Bulletin České statistické společnosti* **21**(3), 89–94.

- [8] Ritov Y. (1990): Estimation in a linear regression model with censored data. *Annals of Statistics* **18**, 303–328.
- [9] Smith P. J. (2004): Using linear regression techniques with censored data. *Int. J. Reliab. Qual. Safety Eng.* **11**(2), 163–173.
- [10] Tsiatis A. A. (1990): Estimating regression parameters using linear rank tests for censored data. *Annals of Statistics* **18**, 354–372.

VZPOMÍNKY DOC. JANA ŘEHÁKA NA DOC. JIŘÍHO ŽVÁČKA

JAN ŘEHÁK'S MEMORIES ON JIŘÍ ŽVÁČEK

Jan Řehák

Je to již více než rok, co 24. listopadu 2015 zemřel doc. Jiří Žváček. V dosud publikovaných vzpomínkách nebyla zdůrazněna charakteristika Jiřího v jeho smyslu pro humor, někdy až sarkastický, ale vždy mířený pozitivně a s podporou kladného průběhu věcí.

Jirka měl těžký osud, ale nesl jej statečně a pokorně a ani na lůžku nebyl pasivní a nikdy nepřestal být svůj. Pracoval a angažoval se pro dobro podobně tragicky postižených. A nepřestával se zajímat o statistiku a o nové ve statistickém softwaru. Udržovali jsme s ním kontakt po celou dobu a poskytovali jsme mu nejnovější verze našeho softwaru. Svou silnou vůlí nikdy neztratil kontakt s vývojem naší vědy. V tom jsem ho obdivoval. V mnohém se naše názory lišily, ale vždy naše debaty byly věcné a plné jasných argumentů. Myslím, že nejen já, ale i Jirka byl našimi výměnami argumentů obohacen. Rád na to vzpomínám.

Byl to dobrý kolega a kamarád. Spolehlivý. Náročný a vyžadující. Mé návštěvy u Jiřího v nemocnici mně vždy ukázaly, jak tenká je hranice mezi standardním životem a tragédií a jak moc si musíme vážit schopnosti pohybu a toho, že se nám neštěstí prozatím vyhnulo. Na Jirku často vzpomínám, protože on byl výraznou osobností v naší profesi a silnou osobností ve svém údělu.

ODEŠEL PROF. RNDR. ING. IVAN KŘIVÝ, CSC. PROFESSOR IVAN KŘIVÝ PASSED AWAY

Josef Tvrdík

V říjnu 2016 zemřel po statečném souboji s rakovinou prof. RNDr. Ing. Ivan Křivý, CSc., dlouholetý člen České statistické společnosti, aktivní účastník řady akcí společnosti (např. ROBUST, STAKAN, Statistické dny) a několika konferencí COMPSTAT.

Ivan Křivý se narodil v roce 1940 na Valašsku a jeho rodiště ho vybavilo velikou odolností a vitalitou. Po gymnáziu vystudoval v letech 1957–1962 Fakultu technické a jaderné fyziky ČVUT a pak při zaměstnání v Ústavu jaderného výzkumu v Řeži i fyziku na MFF UK. V Ústavu jaderného výzkumu se úspěšně zabýval krystalografií a z té doby pochází několik jeho dodnes citovaných článků v mezinárodních časopisech. Uprostřed sedmdesátých let pracoval v PVT v Holešově a v roce 1977 nastoupil na katedru matematiky Pedagogické fakulty v Ostravě, kde učil především pravděpodobnost a matematickou statistiku. Pro tyto předměty napsal i skripta, která stále stojí za přečtení. Na počátku devadesátých let se významně podílel na transformaci Pedagogické fakulty na Ostravskou univerzitu a na její nové Přírodovědecké fakultě byl jejím druhým děkanem. V devadesátých letech přešel na katedru informatiky, kde se kromě výuky zabýval simulačními modely a stochastickými algoritmy.

Měl jsem to štěstí s Ivanem spolupracovat čtvrt století. Podařilo se nám publikovat několik nových algoritmů pro aplikaci evolučních metod v odhadu parametrů nelineárních regresních modelů a ve shlukování. Náš poslední společný článek jsme stihli publikovat v *Applied Soft Computing* koncem roku 2015.

Ivana jsem obdivoval pro jeho hluboké matematické znalosti a schopnost jejich rychlé a účelné aplikace. Myslím, že jsme se při naší práci dobře doplňovali. Ivan občas korigoval mou přehnaně odvážnou rozevlátost, já jsem zase neměl jeho ostych v komunikaci s editory a na konferencích. Vždycky jsem oceňoval jeho zdravě kritické myšlení a naprostou pracovní spolehlivost. Dohodnuté termíny dílčích kroků vždy stihnul, pro moje občasná drobná zpoždění měl pochopení. Společná práce na dokončování finálních verzí článků nám šla od ruky a byly to radostné hodiny. V Ivanovi Křivém jsme ztratili význačného a uznávaného kolegu a mnozí z nás i dobrého kamaráda. Ivane, budeš nám chybět!

ČLENSKÁ SCHŮZE ČESKÉ STATISTICKÉ SPOLEČNOSTI V ROCE 2017

CZECH STATISTICAL SOCIETY IN 2017

Hana Řezanková

Dne 26. ledna 2017 se konala členská schůze České statistické společnosti. Místem konání byla již tradičně zasedací místnost na Českém statistickém úřadu v Praze. Úvodem předsedkyně Společnosti prof. Řezanková přivítala pana ředitele sekce makroekonomických statistik ČSÚ doc. Ing. Jaroslava Sixtu, Ph.D., kterého požádala o úvodní slovo. Poté prof. Řezanková přednesla zprávu o činnosti České statistické společnosti v roce 2016 a Ing. Löster přednesl zprávu o hospodaření za rok 2016 a plán rozpočtu na rok 2017. Se stanoviskem revizorky, doc. Blatné, seznámila přítomné prof. Řezanková. Všechny přednesené zprávy byly členskou schůzí schváleny. Poté byla zahájena zvaná přednáška Mgr. Michaely Šedové, Ph.D. (Institute of Biostatistics and Analyses, Ltd.), s přednáškou na téma *Praxe statistika ve farmaceutickém průmyslu*.

Po skončení přednášky byly vyhlášeny výsledky tajného hlasování. Předsedou České statistické společnosti byl zvolen RNDr. Marek Malý, CSc., ze Státního zdravotního ústavu. Dalšími členy výboru Společnosti byli zvoleni (v abecedním pořadí):

- prof. RNDr. Jaromír Antoch, CSc. (MFF UK Praha)
- RNDr. Jitka Bartošová, Ph.D. (VŠE, FM Jindřichův Hradec)
- RNDr. Marie Budíková, Dr. (PřF MU Brno)
- prof. RNDr. Gejza Dohnal, CSc. (FS ČVUT v Praze)
- RNDr. Jan Klaschka, Ph.D. (ÚI AV ČR, Praha)
- Ing. Martina Litschmannová, Ph.D. (FEI VŠB-TU Ostrava)
- Ing. Tomáš Löster, Ph.D. (VŠE Praha)
- prof. RNDr. Jan Pícek, CSc. (TU Liberec)
- prof. Ing. Hana Řezanková, CSc. (VŠE Praha)
- Mgr. Ondřej Vencálek, Ph.D. (PřF UP Olomouc)
- Ing. Ondřej Vozár (ČSÚ)
- RNDr. Marta Žambochová, Ph.D. (UJEP Ústí nad Labem)

Revizorkou byla zvolena doc. Ing. Dagmar Blatná, CSc. (VŠE Praha).

Na první schůzi výboru České statistické společnosti, která se konala po skončení členské schůze, byli zvoleni další členové předsednictva, které bude pracovat ve složení:

- RNDr. Malý – předseda,
- prof. Řezanková – místopředsedkyně,
- Ing. Litschmannová – vědecká tajemnice,
- Ing. Löster – hospodář.

Dále se výbor usnesl, aby další důležité funkce vykonávali členové výboru a ČStS jako doposud, tj. webové stránky bude i nadále spravovat Ing. Litschmannová, šéfredaktorem časopisu *Informační bulletin České statistické společnosti* zůstává prof. Dohnal a odpovědným redaktorem Mgr. Vencálek. Činnost technického redaktora bude i nadále vykonávat Ing. Pavel Stríž.

Zapsáno v Praze, 26. ledna 2017.

POZVÁNKY NA KONFERENCE CONFERENCE INVITATIONS

Redakce

MOVISS: Bio and Data

Ve dnech 20.–23. září 2017 se v rakouském Vorau uskuteční konference MOVISS určená pro bioinformatiky, biochemiky, statistiky a všechny ty, kteří zpracovávají a interpretují data z oblasti metabolomiky.

Nejde o klasickou konferenci, na níž jsou prezentovány pouze dosažené výsledky. Cílem konference MOVISS je rovněž vyvolání diskuze zaměřené na identifikaci a řešení současných problémů týkajících se nakládání s daty z oblasti metabolomiky.

Více informací naleznete na www.MOVISS.eu.

PRASTAN 2017

Ve dnech 5. října (čtvrtek) – 8. října (neděle) 2017 se na chatě Šípková (<http://sipkova.sk/>), která se nachází 5 km od centra obce Terchová na Slovensku, uskuteční další ročník slovensko-české konference PRASTAN. Tato konference je společnou akcí Slovenskej štatistickej a demografickej spoločnosti

a České statistické společnosti. Tyto společnosti organizují česko-slovenské konference každý lichý rok a země se pravidelně střídají. V České republice se konference koná pod názvem STAKAN a na Slovensku pod názvem PRASTAN.

V letošním roce je tato akce pořádána ve spolupráci s Katedrou matematiky a deskriptivní geometrie ze Stavební fakulty Slovenské technické univerzity v Bratislavě. PRASTAN 2017 bude věnován pravděpodobnosti, statistice, numerické matematice a jejím aplikacím.

Více informací naleznete <http://www.math.sk/wiki/abstrakty17>.

ROBUST 2018

Ve dnech 21. ledna (neděle) – 26. ledna (pátek) 2018 se v rekreačním zařízení Rybník na rozmezí Šumavy a Černého lesa uskuteční jubilejí dvacátá letní škola JČMF ROBUST 2018. Tato akce je organizována skupinou pro výpočetní statistiku ČMS JČMF za podpory ČStS a KPMS MFF UK. ROBUST 2018 bude věnován, tak jako vždy, vybraným trendům matematické statistiky, finanční matematiky, optimalizace, teorie pravděpodobnosti a analýzy dat.

Vedle zvaných přednášek se samozřejmě počítá se sděleními účastníků; pro doktorandy a studenty připravujeme opět posterovou sekci spojenou s ústní prezentací. Jeden večer bude věnován diskusi o nových trendech ve výpočetním prostředí pro analýzu dat a vystoupení zástupců firem, které je šíří.

Více informací naleznete v <http://www.statspol.cz/cs/wp-content/uploads/robust2018.pdf>.

Registrace: od 3. 3. 2017 do 30. 9. 2017, <https://robust.nipax.cz/>. Pokud budete mít jakékoliv potíže s registrací, obraťte se, prosím, na kolegu Dohnala na emailové adrese dohnal@nipax.cz.

Místo konání: Rekreační zařízení Rybník, www.rybnik-ceskyles.cz: GPS 49°30'80,265"N, 12°40'65,457"E.

Novinky hledejte na webové stránce profesora Antocha: <http://www.karlin.mff.cuni.cz/~antoch/>.

Obsah

Vědecké a odborné články

Petr Volf

Model konkurenčních rizik s netradiční aplikací 1

Petr Novák

Diagnostické metody pro model zrychleného času 12

Zprávy a informace

Jan Řehák

Vzpomínky doc. Jana Řeháka na doc. Jiřího Žváčka 20

Josef Tvrdík

Odešel prof. RNDr. Ing. Ivan Křivý, CSc. 21

Hana Řezanková

Členská schůze České statistické společnosti v roce 2017 22

Redakce

Pozvánky na konference 23

Informační bulletin České statistické společnosti vychází čtyřikrát do roka v českém vydání. Příležitostně i mimořádné české a anglické číslo. Vydavatelem je Česká statistická společnost, IČ 00550795, adresa společnosti je Na padesátém 81, 100 82 Praha 10. Evidenční číslo registrace vedené Ministerstvem kultury ČR dle zákona č. 46/2000 Sb. je E 21214. Časopis je na Seznamu recenzovaných neimpaktovaných periodik vydávaných v ČR, více viz server <http://www.vyzkum.cz/>.

The Information Bulletin of the Czech Statistical Society is published quarterly. The contributions in the journal are published in English, Czech and Slovak languages.

Předseda společnosti: RNDr. Marek MALÝ, Státní zdravotní ústav, Šrobárova 48, Praha 10, 100 42, e-mail: mmaly@szu.cz.

Redakce: prof. RNDr. Gejza DOHNAL, CSc. (šéfredaktor), prof. RNDr. Jaromír ANTOCH, CSc., prof. Ing. Václav ČERMÁK, DrSc., doc. Ing. Jozef CHAJDIK, CSc., doc. RNDr. Zdeněk KARPÍŠEK, CSc., RNDr. Marek MALÝ, CSc., doc. RNDr. Jiří MICHÁLEK, CSc., prof. Ing. Jiří MILITKÝ, CSc., doc. Ing. Iveta STANKOVIČOVÁ, Ph.D., doc. Ing. Josef TVRDÍK, CSc., Mgr. Ondřej VENCÁLEK, Ph.D.

Redaktor časopisu: Mgr. Ondřej VENCÁLEK, Ph.D., ondrej.vencalek@upol.cz.
Informace pro autory jsou na stránkách společnosti, <http://www.statspol.cz/>.

DOI: 10.5300/IB, <http://dx.doi.org/10.5300/IB>
ISSN 1210–8022 (Print), ISSN 1804–8617 (Online)

Toto číslo bylo vtištěno s laskavou podporou Českého statistického úřadu.