MARTINGALINIŲ ĮVERČIŲ TAIKYMAS IŠLIKIMO TEORIJOJE

Erika Sabutytė, Vaidotas Kanišauskas
Šiaulių universitetas, Matematikos ir informatikos fakultetas

Įvadas

Martingalinai įverčiai išlikimo teorijoje buvo pradėti taikyti praeito amžiaus aštuntajame dešimtmetyje, pasirodžius O. Alano straipsniui (1978 m.). Po to publikuota daug šios srities darbų, galima paminėti [2], [3]. Mūsų darbe nagrinėjamas skaičiuojantis procesas, kurio rizikos greičio funkcija turi gana sudėtingą pavidalą. Gautas ieškomos funkcijos martingalinis įvertis, įrodomas to įverčio tolydus pagrįstumas ir asimptotinis normalumas.

Tarkime, kad tikimybinėje erdvėje (Ω, F, P) apibrėžtas skaičiuojantis procesas:

$$\boldsymbol{N_n(t)} = \sum_{i=1}^{n} 1(\boldsymbol{X_i \le t}) \quad t \in [0,+\infty),$$

kuriame atsitiktiniai dydžiai $X_1, X_2, \ldots, X_n$ yra nepriklausomi, neneigiami ir absoliučiai tolydūs. Juos charakterizuoja pasiskirstymo arba rizikos greičio funkcija:

$$\alpha_i(t) = \frac{f(t)}{1 - F_i(t)},$$

kur $F_i(t) = P(X_i \le t)$, tankis $F`_i(t) = f_i(t)$.

Mūsų tikslas – išnagrinėti skaičiuojantį procesą, kurio atsitiktiniai dydžiai X_i turi

$$\alpha_i(t) = k_i(t)[\alpha(t) + \mu_i(t)],\ i = \overline{1,n} \tag{1}$$

čia α(s) – nežinoma funkcija, o $\mu_i(s)$ ir $k_i(s)$ žinoma rizikos greičio funkcijos.

Mums žinomuose šaltiniuose tokio pavidalo (1) funkcija dar nebuvo nagrinėta.

1. Lema [6]: Tarkime, kad turime nepriklausomus atsitiktinius dydžius $X_1, X_2, \ldots, X_n$, kuriems galioja riba:

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} DX_i = 0,$$

tada dalinė suma $S_n = X_1 + \ldots + X_n$ tenkina didžiųjų skaičių tokio pavidalo dėsnį:

$$\left| \frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} \mathsf{E} X_i \right| \xrightarrow{P} 0,\ n \to \infty.$$

čia E – žymi matematinį vidurkį.

2. Lema [4]: Jei M(t) yra kvadratu integruojamas lokalus martingalas, su visais $\delta > 0$ ir $\eta > 0$, $t \in [0,+\infty)$ tenkina nelygybę:

$$P(\sup_{s\in[0,t]} |M(t)| > \eta) \le \frac{\delta}{\eta^2} + P(\langle M(t)\rangle > \delta). \tag{2}$$

Paprastai nežinoma funkcija α(*t*) nėra vertinama, tik vertinamas jos integralas:

$$L(t) = \int_0^t \alpha(s)ds.$$

Pirmiausia rasime šios funkcijos martingalinį įvertį. Tam tikslui randame skaičiuojančio proceso $N_n(t)$ natūralų kompensatorių. Kaip žinome [2], jis yra tokio pavidalo:

$$A_n(t) = \sum_{i=1}^{n} \int_0^t \alpha_i(s) Y_i(s) ds, \tag{3}$$

čia $\boldsymbol{Y(s)} = 1(\boldsymbol{X_i \ge s})$, o $1(\boldsymbol{A})$ yra aibės A indikatorius.

Įstatę į šią formulę (1), gauname:

$$A_n(t) = \sum_{i=1}^{n} \int_0^t k_i(s)(\alpha_i(t) + \mu_i(s)) \cdot Y_i(s) ds,$$

kurią pertvarkę gauname:

$$A_n(s) = \int_0^t \alpha(s) H_n(s) ds + \int_0^t K_n(s) ds,$$

kur $\boldsymbol{H_n(s)} = \sum_{i=1}^{n} \boldsymbol{k_i(s) Y_i(s)} = \sum_{i=1}^{n} \boldsymbol{k_i(s)} 1(\boldsymbol{X_i \ge s})$

$$\boldsymbol{K_n(s)} = \sum_{i=1}^{n} \boldsymbol{k_i(s)} \mu_i(\boldsymbol{s}) \boldsymbol{Y_i(s)} = \sum_{i=1}^{n} \boldsymbol{k_i(s)} \mu_i(\boldsymbol{s}) 1(\boldsymbol{X_i \ge s}), \tag{4}$$

Dėl (4) formulėje esančių pointegralinių funkcijų absoliutaus tolydumo galima (4) formulę pertvarkyti į pavidalą:

$$L(t) = \int_0^t \alpha(s) ds = \int_0^t \frac{dA_n(s)}{H_n(s)} - \int_0^t \frac{K_n(s)}{H_n(s)} ds.$$

Dabar *L*(*t*) martingalinis įvertis gaunamas vietoje kompensatoriaus $A_n(s)$ įrašius taškinį procesą $N_n(s)$:

$$\hat{L}(t) = \int_0^t \frac{dN_n(s)}{H_n(s)} - \int_0^t \frac{K_n(s)}{H_n(s)} ds,$$

gavome lokalųjį martingalą:

$$\hat{L}_n(t) - L(t) = \int_0^t \frac{dM_n(s)}{H_n(s)} = \int_0^t \frac{d(N_n(s) - A_n(s))}{H_n(s)},$$

kaip žinome [4], lokalaus martinglo vidurkis lygus nuliui: $\mathsf{E}(\hat{L}_n(t) - L(t)) = 0$, vadinasi, $\hat{L}_n(t)$ yra nepaslinktas $L(t)$ įvertis.

Randame to martingalo kvadratinę variaciją:

$$< \hat{L}_n(t) - L_n(t) >= \int_0^t \frac{1}{H^2{}_n(s)} dA_n(s) = \int_0^t \frac{1}{H^2{}_n(s)} (\alpha(s) H_n(s) ds + K_n(s) ds) =$$

$$= \int_0^t \frac{\alpha(s) H_n(s)}{H^2{}_n(s)} ds + \int_0^t \frac{K_n(s)}{H^2{}_n(s)} = \frac{1}{n} \left(\int_0^t \frac{\alpha(s)}{\frac{1}{n} H_n(s)} ds + \int_0^t \frac{\frac{1}{n} K_n(s)}{(\frac{1}{n} H_n(s))^2} ds \right). \tag{5}$$

Įvedame sąlygas:

A: funkcijos $k_i(s)$ ir $\mu_i(s)$, $i = \overline{1,n}$ tokios, kad:

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^n k^2{}_i(s) = 0, \ \lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^n k^2{}_i(s) \mu^2(s) = 0.$$

Egzistuoja funkcijos a(s) ir b(s) tokios, kad su visais $t \in [0,+\infty)$.

$$\inf_{s\in[0,t]} a(s) > 0, \ \inf_{s\in[0,t]} b(s) > 0.$$

$$\lim_{n\to\infty} \frac{1}{n} \mathsf{E} H_n(s) = a(s), \lim_{n\to\infty} \frac{1}{n} \mathsf{E} K_n(s) = b(s). \tag{6}$$

Teorema: Tarkime, kad patenkintos A sąlygos. Tada, kai $n \to \infty$, $t \in [0,+\infty)$.

$$\sup_{s\in[0,t]} \left| \hat{L}_n(t) - L(t) \right| \xrightarrow{P} 0 \tag{7}$$

$$n^{\frac{1}{2}} < \hat{L}_n(t) - L(t) > \Rightarrow N(0, \int_0^t \frac{\alpha(s)}{a(s)} ds + \int_0^t \frac{b(s)}{(a(s))^2} ds \cdot \tag{8}$$

Teoremos įrodymas remiasi (2) nelygybe lokaliems kvadratu integruojamiems martingalams. Mūsų atveju $\hat{L}_n(t) - L(t)$ kaip tik ir yra lokalus kavadratu inegruojamas martingalas. Remiantis (5) formule ir (6) sąlygomis gaunama, kad:

$$\hat{L}_n(t) - L(t) \xrightarrow{P} 0, \ n \to \infty. \tag{9}$$

Vadinasi, pagal (2) nelygybę su kiekvienu $\varepsilon > 0$ teisinga formulė:

$$\lim_{n\to\infty} P(\sup_{s\in[0,t]} \left| \hat{L}_n(t) - L(t) \right| > \varepsilon) = 0,$$

kas yra tapatu (7) formulei.

Kadangi lokalaus martingalo $n^{\frac{1}{2}} \cdot (\hat{L}_n(t) - L(t)$ kvadrato įvertis yra kvadratinė variacija $n < (\hat{L}_n(t) - L(t) >$ ir ji yra baigtinė funkcija, kai n neapibrėžtai didėja dėl (6) sąlygų, kai t fiksuotas, tai jam galima taikyti centrinę ribinę teoremą. Šio martingalo vidurkis lygus nuliui, o dispersija sutampa su (5) formulės dešinės pusės riba, padaugintai iš n, kai $t \to \infty$. Vadinasi:

$$n^{\frac{1}{2}} < \hat{L}_n(t) - L(t) > \Rightarrow N(0, \int_0^t \frac{\alpha(s)}{a(s)} ds + \int_0^t \frac{b(s)}{(a(s))^2} ds,$$

čia rodyklė žymi silpną konvergavimą, o $N(0,\sigma^2)$ – normalųjį tikimybinį dėsnį.

Teorema įrodyta.

Pastaba: Pastebėsime, kad (8) formulė nėra labai patogi fukcijos $L(t) = \int_0^t \alpha(s) ds$ pasikliautinajam intervalui ieškoti, nes į dispersiją įeina nežinoma funkcija α(s). Čia vietoje martingalo $\hat{L}_n(t) - L(t)$ kvadratinės charakteristikos $<\hat{L}_n(t) - L(t)>$ galima vartoti jos martingalinį įvertį – opcionalę charakteristiką $[\hat{L}_n(t) - L(t)]$, kuri asimptotiškai su tikimybe 1 sutampa su $<\hat{L}_n(t) - L(t)>$.

Primename, kad $<\hat{L}_n(t) - L(t)>$ užrašoma ir tokiu pavidalu:

$$< \hat{L}_n(t) - L(t) >= \int_0^t \frac{1}{H_n^2(s)} dA_n(s).$$

Vietoje $A_n(s)$ įrašę taškinį procesą $N_n(t)$, gauname opcionalę charakteristiką:

$$[\hat{L}_n(t) - L(t)] = \int_0^t \frac{1}{H_n^2(s)} dA_n(s).$$

Dabar tpagal centrinę ribinę teoremą martingalams, kai $n \to \infty$, gauname:

$$\frac{\hat{L}_n(t) - L(t)}{([\hat{L}_n(t) - L(t)])^{\frac{1}{2}}} = \frac{\hat{L}_n(t) - L(t)}{(\int_0^t \frac{1}{H_n^2(s)} dN_n(s))^{\frac{1}{2}}} \Rightarrow N(0,1). \tag{10}$$

Ši formulė taikoma L(t) pasikliautinajam intervalui ieškoti:

$$P \left(-u_{1-\frac{\alpha}{2}} < \frac{\hat{L}_n(t) - L(t)}{(\int_0^t \frac{1}{H_n^2(s)} dN_n(s))^{\frac{1}{2}}} < u_{1-\frac{\alpha}{2}} \right) = 1 - \alpha.$$

Pertvarkę skliaustų viduje esančią išraišką, gauname funkcijos L(t) pasikliautiną intervalą:

$$P\left(\hat{L}_n(t)-u_{1-\frac{\alpha}{2}}\cdot\left(\int_0^t\frac{1}{H_n^2(s)}dN_n(s)\right)^{\frac{1}{2}}<L(t)<\hat{L}_n(t)+u_{1-\frac{\alpha}{2}}\cdot\left(\int_0^t\frac{1}{H_n^2(s)}dN_n(s)\right)^{\frac{1}{2}}\right)=1-\alpha .$$

Hipotezių tikrinimas

Tikriname neparametrinę statistinę hipotezę **H_0: L(t) = L_0(t)** esant alternatyvai **H_1: L(t) ≠ L_0(t),** kur $L_0(t)=\int_0^t\alpha_0(s)ds$ – žinoma funkcija.

Jei $L_0(t)\in\left(\hat{L}_n(t)-u_{1-\frac{\alpha}{2}}\cdot\left(\int_0^t\frac{1}{H_n^2(s)}dN_n(s)\right)^{\frac{1}{2}},\hat{L}_n(t)+u_{1-\frac{\alpha}{2}}\cdot\left(\int_0^t\frac{1}{H_n^2(s)}dN_n(s)\right)^{\frac{1}{2}}\right)$, tai gauti duomenys neprieštarauja hipotezės **H_0** teisingumui.

Jei $L_0(t)\notin\left(\hat{L}_n(t)-u_{1-\frac{\alpha}{2}}\cdot\left(\int_0^t\frac{1}{H_n^2(s)}dN_n(s)\right)^{\frac{1}{2}},\hat{L}_n(t)+u_{1-\frac{\alpha}{2}}\cdot\left(\int_0^t\frac{1}{H_n^2(s)}dN_n(s)\right)^{\frac{1}{2}}\right)$, tai **$H_0$** atmetame.

Literatūra

1. Aalen O., 1978, Nonparametric inference for a family of counting proceses, *Ann. Statist*. Nr. 6. P. 701–726.
2. Andersen P. K., Borgan O., Gill R. D., Keiding N., 1993, Statistical Models Base don Counting Processes. New York, Springer-Verlag.
3. Karr A. F. 1986, *Point Processes and their Statistical Inference.* Dekker, New York.
4. Липцер Р. Ш., Ширяев А. Н., 1986, *Теория мартингалов.* Москва: Наука.
5. Канишаускас В., 2002, Асимптотические свойства мартингалных оценок процесов востановления. *Lietuvos matematikos rinkinys* ISSN 0132-2818. Nr. 42, spec. nr. P. 121–126.
6. Kubilius J., 1980, *Tikimybinių teorija ir matematinė statistika.* Vilnius: Mokslas.

APPLICATION OF MARTINGALE ESTIMATORS IN SURVIVAL THEORY

Erika Sabutytė, Vaidotas Kanišauskas

Summary

There is analyzed the complex form of empirical process, $\boldsymbol{N_n(t)=\sum_{i=1}^{n}1(X_i\le t)}$ t ∈ [0,+ ∞), where continuously distributed survival times X_i has hazard rate function $\alpha_i(t) = k_i(t)[\alpha(t) + \mu_i(t)]$, $i=\overline{1,n}$ where the $k_i(t)$ and $\mu_i(t)$ are known hazard rate functions and α(t) is unknown function. In this paper we established the uniform consistency and the asymptotic normality of the martingale estimator of integrated hazard rate $A(t)=\int_0^t\alpha(s)ds$. By using the asymptotic normality of martingale estimators we obtained confidence intervals of *A*(*t*) and by using that tested the corresponding statistical nonparametric hypothesis.

Keywords: martingale estimator, survival theory, survival times, hazard rate function.

MARTINGALINIŲ ĮVERČIŲ TAIKYMAS IŠLIKIMO TEORIJOJE

Erika Sabutytė, Vaidotas Kanišauskas

Santrauka

Nagrinėjamas sudėtingo pavidalo empirinis procesas, kurio rizikos greičio funkcija užrašoma per nežinomos funkcijos sumą į sandaugą su kitomis dviem žinomomis funkcijomis. Gautas nežinomos funkcijos integralo martingalinis įvertis, nustatytos to įverčio asimptotinės savybės: pagrįstumas, nepaslinktumas ir normalumas. Taikant asimptotinį normalumą, gautas ieškomos funkcijos pasikliautinasis intervalas ir patikrinta atitinkama neparametrinė statistinė hipotezė.

Prasminiai žodžiai: martingaliniai įverčiai, išlikimo teorija, atsitiktiniai dydžiai.

Įteikta 2009-09-01