LATVIJAS UNIVERSITĀTE
FIZIKAS UN MATEMĀTIKAS FAKULTĀTE
MATEMATIKAS NODAĻA

IZDZĪVOŠANAS ANALĪZE AR PIELIETOJUMU MEDICĪNISKAJĀ STATISTIKĀ

MAĢISTRA DARBS

Autors: **Olga Grakoviča**

Stud. apl. MaSt020019

Darba vadītājs: doc. Dr.math. Jānis Valeinis

RĪGA 2009

Anotācija

Darba mērķis ir izpētīt izdzīvošanas statistikas teorētisko pamatu, kā arī apskatīt pielietojumu medicīnas datiem. Tiek apskatīts Kaplana-Meijera un Nelsona-Alena izdzīvošanas funkcijas novērtējumi un vairāki veidi ticamības intervālu kostruēšanai. Izdzīvošanas analīzes teorija balstīta uz neparametriskajām un semiparameriskajām statistikas metodēm. Pētīta Koksa proporcionālā riska funkcija. Teorijā aplūkotās metodes pielietotas uz Tuberkulozes un plaušu slimību valsts aģentūras apkopotajiem datiem par 2007.gada pacientiem.

Atslēgas vārdi: izdzīvošanas funkcija, Kaplana-Meijera metode, Nelsona-Alena metode, riska funkcija, Koksa proporcionālā riska funkcija, cenzēti no labās puses

Abstract

The aim of this work is to analyse basic statistical methods in survival analysis. Kaplan-Aalen and Kaplan Meier estimators for survival function have been considered. For these estimators there are sevral ways how to construct confidence intervals. The theory of survival analysis is based on nonparametric and semiparametric models. Finally, Cox proportional hazard model has been analysed. At the end the data of State agency for tuberculosis and lung diseases of Latvia are applied.

Keywords: survival function, Kaplan-Meier model, Nelson-Aalen model, hazard function, Cox proportional harard function

Saturs

Apzīmējumi

$H(t)$ kumulatīvā riska funkcija,

$S(t)$ izdzīvošanas funkcija,

$P(x)$ varbūtība,

$D(x)$ dispersija,

Ievads

Populācijas demogrāfisko situāciju studē sen. Jau 18.gs. bija zināmas pirmās mirstības tabulas, kas ir pirmsākums izdzīvošanas analīzei. Šāda veida tabulas izmanto vēl tagad. Izdzīvošanas analīze ir svarīga medicīnas statistikas pielietojumos. Tomēr jebkurš process, kuram iespējams definēt izdzīvošanu kādā periodā, var tikt raksturots ar šeit apskatītajām metodēm. Tā, piemēram, datortehnikas kalpošanas ilgums var tikt aprēķināts izmantojot izdzīvošanas analīzi. Parasti interesē izdzīvošanas iespējas dažādu faktoru ietekmē - dzimuma, vecuma, terapijas veida un daudzu citu.

Pētījuma mērķis - izanalizēt izdzīvošanas statistikas novērtējumu metodes un pielietot to uz reāliem datiem. Dati ir novērojumi, kas veikti konkrētā laikā. Dati var būt cenzēti. Metodes balstītas uz Kaplana-Meijera un Nelsona-Alena novērtējumiem, apskatot neparametrisku gadījumu. Kaplana-Meijera ir vienkāršākā datu aprakstīšanas metode. Tā nosaka pacienta varbūtību izdzīvot. Nelsona-Alena novērtējums ir riska funkcija (hazard function). Šis novērtējums ir izdzīvošanas funkcijas pamatā un svarīgs lielums neparametriskās statistikas aprēķinos. Pielietojums matemātiskajam modelim ir pacientu datu apstrāde. Nepieciešama grupa ar cilvēkiem, kuriem konstatēta slimība ar iespējamu nāves gadījuma rezultātu. Par šādu slimību tika izvēlēta tuberkuloze, datus par slimniekiem iegūstot no Tuberkulozes un plaušu slimību valsts aģentūras. Lai šos datus pētītu, jāzina personu slimības konstatēšanas laiks, vecums tajā brīdī, novērojuma laikam jābūt vismaz 1 mēnesim. Cilvēku grupā jāatrodas personām, kas jau ir mirušas no minētās slimības. Tā kā statistika paredzēta cenzētiem datiem, tad iespējami gadījumi, kad ir pazaudēts kontakts ar pacientu. Pētījumā tiek iekļauti 2007. gadā reģistrētie pirmreizēji tuberkulozes pacienti.

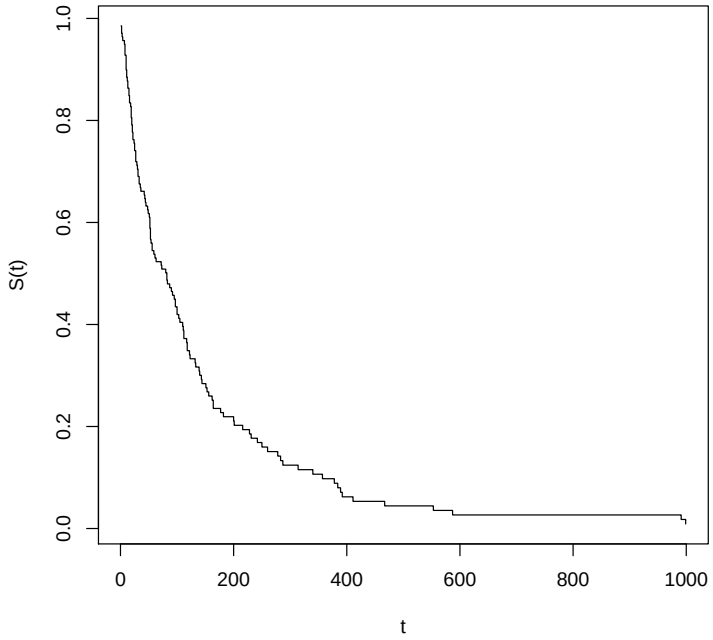
Darbs sastāv no 4.nodaļām. Pirmajā daļā tiek apskatīts datu aprakstošās metodes, otrajā - neparametriski novērtējumi Kaplana-Meijera un Nelsona-Alena novērtējumiem. Trešajā daļā datu salīdzināšanas metodes un pēdējā daļā - metožu pielietojums uz Latvijas tuberkulozes pacientu datiem. Paralēli teorijai tiek konstruēti piemēri.

1. Izdzīvošanas datu attēlošanas metodes

Dati ir novērojumi, kas veikti konkrētā laikā. Novērojums var beigties ar to, ka mēs skaidri zinām, kas ar objektu notiek, t.i. objekts nomirst. Tāpat var būt, ka objekts vairs netiek novērots un nav zināms, vai pēc laika t tas joprojām ir dzīvs. Šādus datus, kur par visiem novērojumiem nav pilnīgas informācijas sauc par **cenzētiem**. Var izdalīt datus cenzētus no labās un no kreisās puses. Pirmajā gadījumā nav zināma informācija par objekta stāvokli pēc laika t, bet otrajā - informācija novērojumu (piem., slimības) sākumu. Turpmāk apskatīsim datus cenzētus no labās puses.

Prakstiski līdzās teorijai, apskatīsim R iebūvēto izdzīvošans datu *veteran* un *pbc* statistikas aprēķinu pielietojumā. Fails *veteran* satur informāciju par plaušu vēža slimniekiem, kuriem izmantotas divas dažādas terapijas. Datu apjoms ir 137 ieraksti. Otrs fails satur ziņas par retas aknu slimības pacientiem, ar 418 ierakstiem. Abos failos dati cenzēti no labās puses.

Pirmajam ieskatam izdzīvošanas analīzē sākumā pamatā tiek izmantots Hosmera (Hosmer) un Lemeshova (Lemeshow) darbs, kas balstās uz metožu pielietojumu,[1].
Tiek pieņemts, ka dati ir nepārtraukti un cenzēti no labās puses. Visa analīze ir balstīta uz kumulatīvas sadalījuma funkcijas novērtēšanu. Būtībā šī sadalījuma funkcija ir varbūtība, ka nejauši izvēlēta pētāmā objekta izdzīvošanas laiks ir mazāks par kādu noteiktu t. Tas tiek definēts kā $F(t) = P(T < t)$. Savukārt **izdzīvošanas funkcija** ir varbūtība izdzīvot kādā laika momentā jeb varbūtība, ka kāds novērtēts laiks būs lielāks par noteikto laiku t, t.i. $S(t) = P(T \geq t)$.



1.1. att. Kaplana-Meijera izdzīvošanas funkcijas novērtējums datiem *veteran*

1.1. Kaplana-Meijera novērtējums

Tiek novērots objektu kopums laika intervālā I. Katra pētāmā objekta novērojuma laiks tiek apzīmēts ar t_i. Pirmais interesējošais laika intervāls ir no $t = 0$ līdz $t = min(t_i)$: $I_0 = (t : 0 \leq t < min(t_i)$. Nākamais intervāls ir no minimālā novērojuma līdz nākamajam mazākajam novērojuma laikam un tā tālāk.

Laikā no $t = 0$ līdz minimālajam novērojuma laikam novērtētā izdzīvošanas varbūtība visiem pētāmajiem objektiem ir vienāda ar 1, $\hat{S}(t) = 1$. Turpmāk ar n apzīmējam novērojumu skaitu (izlases lielumu), bet ar d nāves gadījumu skaitu. Visas turpmākās varbūtības katrā intervālā tiek izteiktas ar formulu

$$\hat{S}(t) = \prod_{t_i \leq t} \frac{n_i - d_i}{n_i} \tag{1.1.1}$$

ar pieņēmumu, ka $\hat{S}(t) = 1$, ja $t < t_i$. 1.1. attēls parāda šīs funkcijas atkarību no laika datiem *veteran*.

Sākumā augstā līkne liek secināt, ka daudzi no pacientiem nomira drīz pēc novērošanas sākuma. Salīdzinoši garā labā aste veidojas, jo ir daži cilvēki, kuriem bijis garš dzīves laiks. Līknes veids atkarīgs no novērtētajiem izdzīvošanas laikiem un novērtējumiem, kas ir cenzēti.

Izdzīvošanas funkcija pareizi attēlotu izdzīvošanas iespējas, ja dati nebūtu cenzēti. Tā kā nav zināms, vai cilvēki, par kuriem novērojumi pārtrūka, izdzīvo līdz mums interesējošam laikam t, tad izdzīvošanas funkcija jākoriģē, ņemot vērā, ka kāda daļa no tiem nomirst. Mainīgo, kas parāda, vai dati ir cenzēti, apzīmē ar c. Tas pieņem divas vērtības - 1, ja iestājusies nāve, 0 - informācija par notikumu nav zināma. Pieņemot, ka cenzētie novērojumi ir sadalīti vienmērīgi, izdzīvošanas funkciju var modificēt uz šādu izteiksmi:

$$\hat{S}(t) = \prod_{t_i \leq t} \frac{n_i - (c_i/2) - d_i}{n_i - (c_i/2)} \tag{1.1.2}$$

Alternatīva Kaplana-Meijera novērtējumiem ir mirstības tabulas, kas attēlo izdzīvošanu konkrētos laika intervālos. Kaplana-Meijera metode attēlo izdzīvošanas funkciju tieši no nepārtrauktas izdzīvošanas vai miršanas notikumiem, t.i. tā nav atkarīga no konkrētiem laika intervāliem.

Lai konstruētu ticamības intervālus, iegūsim dispersijas novērtējumu. Ir vairāki veidi, kā iespējams iegūt Kaplana-Meijera novērtējuma dispersiju. Aplūkosim tā saucamo *delta metodi*, kuras pamatā ir Teilora rindas izvirzījums. Novērtējums ir attiecību reizinājums, taču vieglāk strādāt ar summu, pirms tam veicot logaritmisku transformāciju. Pierādījumu, šim novērtējumam ir 2. pielikumā. Rezultātā Kaplana-Meijera novērtējums izskatās kā attiecību summa

$$\ln(\hat{S}(t)) = \sum_{t_i \leq t} \frac{n_i - d_i}{n_i} = \sum_{t_i \leq t} \hat{p}_i, \tag{1.1.3}$$

kur $\hat{p}_i = (n_i - d_i)/n_i$. Ja uzskatām, ka novērojumi laikā t_i ir neatkarīgi Bernulli sadalīti ar konstantu varbūtību, tad $\hat{p}_i$ ir varbūtības novērtējums un tā dispersijas novērtējums ir $(\hat{p}_i(1 - \hat{p}_i))/n_i$. Izmantojot delta metodi, dispersijas novērtējums mainīgajam X ir aptuveni

$$D[\ln(X)] \cong \frac{1}{\mu_X^2}\sigma_X^2,$$

kur μ_X ir parametra X vidējais, bet σ_X^2 dispersija. Abu parametra vietā ievieto to novērtējumus. Attiecīgi dispersijas novērtējums logaritmam no novērtētās varbūtības $ln(\hat{p}_i)$ ir

$$\hat{D}[\ln(\hat{p}_i)] \cong \frac{1}{\hat{p}_i^2}\frac{\hat{p}_i(1 - \hat{p}_i)}{n_i} \cong \frac{d_i}{n_i(n_i - d_i)}$$

Ja pieņemam, ka novērojumi ir neatkarīgi, tad izdzīvošanas funkcijas dispersijas novērtējuma logaritmiska transformācija ir uzrakstāma

$$\hat{D}[\ln(\hat{S}(t))] = \sum_{t_i \leq t} \hat{D}[\ln(\hat{p}_i] = \sum_{t_i \leq t} \frac{d_i}{n_i(n_i - d_i)} \tag{1.1.4}$$

Dispersijas novērtējums var izteikt arī caur eksponenciālu transformāciju

$$D(e^X) \cong (e^{\mu_X})^2 \sigma_X^2. \tag{1.1.5}$$

Ņemot vērā, ka $\hat{S}(t) = e^{\ln(\hat{S}(t))}$, X vietā ievieto $\ln(\hat{S}(t))$, σ_X^2 ir dispersija formulā (1.1.4) un μ_X (1.1.5) formula aproksimēts ar $\ln(\hat{S}(t))$, tad dispersijas novērtējumu var uzrakstīt

$$\hat{D}(\hat{S}(t)) = (\hat{S}(t))^2 \sum_{t_i \leq t} \frac{d_i}{n_i(n_i - d_i)}. \tag{1.1.6}$$

Tagad varam konstruēt punktveida ticamības intervālu novērtējumu. Tiek pieņemts, ka Kaplana-Meijera novērtējums ir asimptotiski normāli sadalīts, kas no vienas puses ļautu viegli aprēķināt ticamības intervālu, izdzīvošanas funkcijai atņemot vai pieskaitot standartkļūdas reizinājumu ar standarta normāla sadalījuma kvantili, no otras puses aprēķini tiek ierobežoti, ja izlase nav pietiekami liela. Lai atrisinātu šo problēmu, tiek ieteikts ticamības intervāla konstruēšanai par pamatu ņemt

$$\ln[-\ln(\hat{S}(t))],$$

ko sauc par log-log izdzīvošanas funkciju. Tās dispersijas novērtējums ir

$$\hat{D}(\ln[-\ln(\hat{S}(t))]) = \frac{1}{[\ln(\hat{S}(t))]^2} \sum_{t_i \leq t} \frac{d_i}{n_i(n_i - d_i)}. \tag{1.1.7}$$

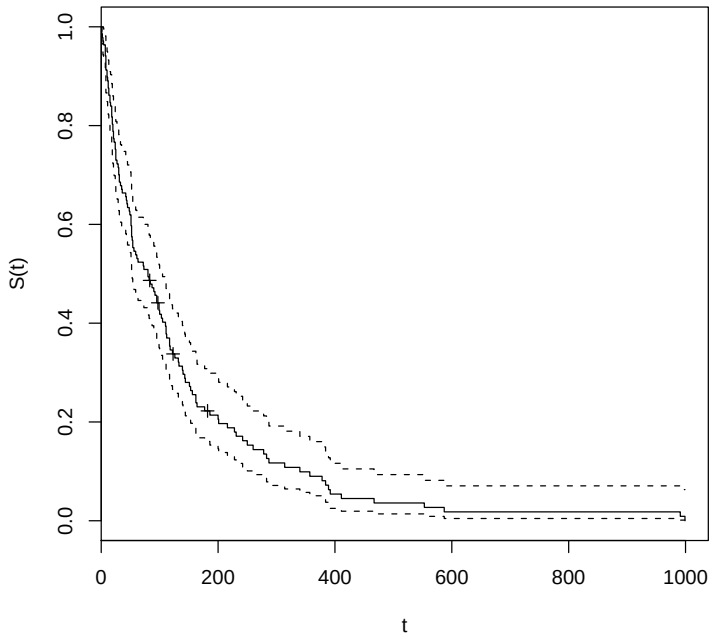
Šo izteiksmi sauc par Grīnvuda (Greenwood) formulu.

$100(1-\alpha)$ procentīga ticamības intervāla galapunkti log-log izdzīvošanas funkcijai ir

$$(\ln[-\ln(\hat{S}(t))]) \pm z_{\alpha/2} \hat{D}(\ln[-\ln(\hat{S}(t))]), \tag{1.1.8}$$

kur $z_{\alpha/2}$ standartnormāla sadalījuma augšējā $\alpha/2$ kvantile un $\hat{SE}(.)$ ir novērtētā standartkļūda, kas šajā gadījumā ir kvadrātsakne no (1.1.7). Ja augšējo un apakšējo ticamības intervāla robežu attiecīgi apzīmē ar $\hat{c}_l$ un $\hat{c}_u$, tad izteiksmi (1.1.8) var pārrakstīt formā

$$\exp[-\exp(\hat{c}_u)] \; un \; \exp[-\exp(\hat{c}_l)] \tag{1.1.9}$$



1.2. att.: Kaplana-Meijera izdzīvošanas funkcijas novērtējums un tā punktveida ticamības joslas, datiem *veteran*

1.2. attēlā redzam ticamības intervālu izdzīvošanas funkcijas novērtējumam datiem *veteran*.

Halls (Hall) un Velners (Wellner) prezentē vienlaicīgos ticamības intervālus. Tie parasti nav tik izplatīti, kā punktveida ticamības intervāli, jo nepieciešamsa zināt izdzīvošanas funkcijas sadalījuma kritiskās vērtības. 1.pielikumā ir Halla un Velnera iegūtā kritisko vērtību tabula. Lai intervālus varētu salīdzināt ar punktveida robežām, arī tiem aprēķinus veiksim transformētai izdzīvošanas funkcijai. Intervālus iesaka ierobežot ar laikiem, kas ir mazāki vai vienādi ar lielāko novērtēto izdzīvošanas laiku, t.i., lielāko necenzēto laiku, kas tiek apzīmēts ar t_m. Pie $100(1-\alpha)$ procentīgas ticamības robežas intervālam $[0, t_m]$ log-log izdzīvošanas funkcijas transformācija ir

$$(\ln[-\ln(\hat{S}(t))]) \pm H_{\hat{\alpha},\alpha} \frac{1 + n\hat{\sigma}^2(t)}{\sqrt{n}|\ln(\hat{S}(t))|},$$

kur

$$\hat{\sigma}^2(t) = \sum_{t_i \leq t} \frac{d_i}{n_i(n_i - d_i)},$$

Kaplana-Meijera novērtējuma formulā (1.1.4) dispersijas novērtējums un $H_{\hat{\alpha},\alpha}$ ir kritiskā vērtība no tabulas 1.pielikumā, kur

$$\hat{\alpha} = n\hat{\sigma}^2(t_m)/[1 + n\hat{\sigma}^2(t_m)].$$

Un atkal - ja zemāko robežu apzīmē ar $\hat{b}_l$ un augstāko ar $\hat{b}_u$, tad ticamības robežas izdzīvošanas funkcijai ir

$$\exp[-\exp(\hat{b}_u)] \, un \, \exp[-\exp(\hat{b}_l)] \tag{1.1.10}$$

1.2. Nelsona-Alena novērtējums

Kaplana-Meijera novērtējums ir praksē visbiežāk izmantotais novērtējums, kurš atrodams vairumā programmatūru. Tagad apskatīsim vēl vienu izdzīvošanas funkcijas novērtējumu. Pieņemam, ka laika mainīgais ir nepārtraukts, tad izdzīvošanas funkciju var uzrakstīt kā

$$S(t) = e^{-H(t)}, \tag{1.2.1}$$

kur $H(t) = -\ln(S(t))$. Formula (1.2.1) parāda, ka izdzīvošanas funkcijas pamatā var būt S(t) - Kaplana-Meijera novērtējums vai novērtējums $H(t)$. Funkciju $H(t)$ sauc par **Nelsona-Alena** (Nelson-Aalen) novērtējumu. To izsaka funkcija:

$$\hat{H}(t) = \sum_{e}^{t_i \leq t} \frac{d_i}{n_i}. \tag{1.2.2}$$

Kā redzam - $H(t)$ ir apgriezts jēdziens izdzīvošanas funkcijai, t.i. raksturo varbūtību nomirt, nevis izdzīvot. 1.3. grafikā attēlots $\hat{H}(t)$ novērtējums *veteran* datiem.

Novērtējums izdzīvošanas funkcijai šajā gadījumā ir

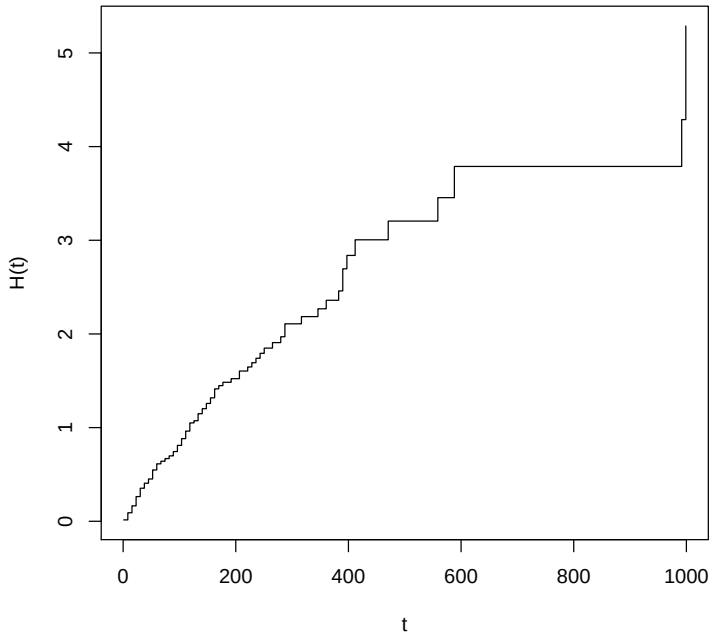
$$\hat{S}(t) = e^{-\hat{H}(t)}. \tag{1.2.3}$$

1.4. attēlā var salīdzināt Kaplana-Meijera un Nelsona-Alena izdzīvošanas funkcijas novērtējumus piemēra datiem. Nelsona-Alena metode izdzīvošanas funkcijas konstruēšanai vienmēr dod lielāku novērtējumu nekā Kaplana-Meijera.

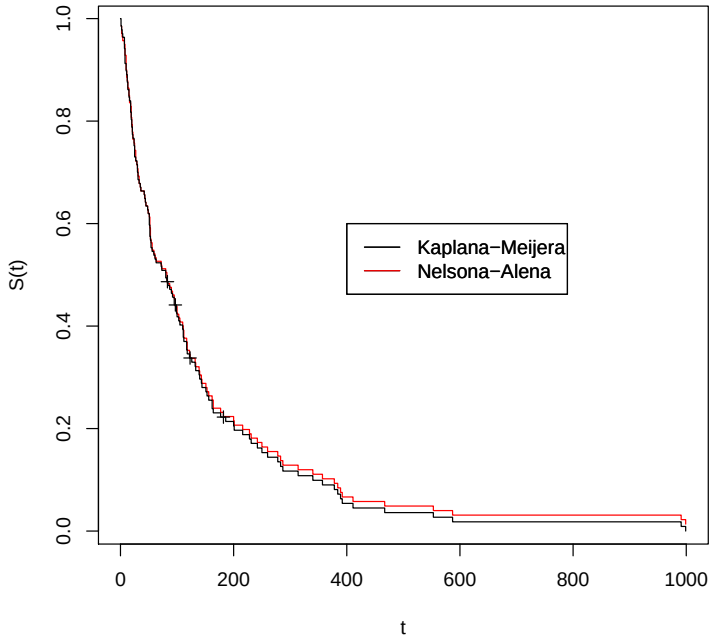
1.2.1. Riska funkcija

Augstāk apskatītā funkcija $H(t)$ ir svarīgs izdzīvošanas analīzes objekts. Šo funkciju sauc par **kumulatīvo riska funkciju** (hazard function).

Šis termins lietots, lai apzīmētu risku, ar kādu intervālā pēc laika t nevēlamais notikums notiek, balstoties uz informāciju par pētāmā objekta izdzīvošanu līdz laikam t.



1.3. att. Nelsona-Alena riska funkcijas novērtējums $\hat{H}(t)$ datiem *veteran*



1.4. att.: Kaplana-Meijera izdzīvošanas funkcijas novērtējums salīdzinājumā ar Nelsona-Alena novērtējumu datiem *veteran*

Riska funkcijai ir liela nozīme izdzīvošanas regresiju analīzē.

Apskatīsim detalizētāk riska funkcijas matemātisko konstrukciju. Iepriekš tika definēta izdzīvošanas funkcija ka $S(t) = P(T \geq t)$. Parasti mūs interesē gadījumi, kad ir nepārtraukts sadalījums ar blīvuma funkciju:

$$f(t) = -S_{\prime}(t) = \lim_{\delta \to 0+} \frac{P(t \leq T < t + \delta)}{\delta}, \tag{1.2.4}$$

līdz ar to izdzīvošanas funkciju var uzrakstīt

$$S(t) = \int_t^{\infty} f(u)du.$$

Diskrētam gadījuma lielumam sadalījumi parasti tiek izteikti, izmantojot varbūtību blīvuma komponenti $f_j\sigma(t - a_j)$ elementam f_j laikā a_j, kur $\sigma(.)$ ir Diraka (Dirac) delta funkcija. Diraka delta funkcija raksturo bezgalīgi augstu pīķu robežu apgabalu. Funkcijai visu argumentu apgabalā vērtība ir 0, izņemot gadījumu, kad arguments vienāds ar 0, tad funkcijas vērtība ir bezgalīgi liela:

$$\sigma(x) = \begin{cases} +\infty, x = 0 \\ 0, x \neq 0 \end{cases}$$

Šīs funkcijas integrālis vienāds ar 1: $\int_{-\infty}^{\infty} \sigma(x)dx = 1$.

Funkcijas $F_T(.)$ un $f_T(.)$ ir divi matemātiski ekvivalenti veidi kā uzdot nepārtraukta, nenegatīva gadījuma lieluma sadalījumu. Definēsim vēl vienu sadalījuma raksturotāju - riska funkciju:

$$h(t) = lim_{\delta \to 0+} \frac{O(t \leq T < t + \delta | t \leq T)}{\delta}. \tag{1.2.5}$$

Pēc nosacītās varbūtības definīcijas iegūstam

$$h(t) = f(t)/S(t).$$

Diskrētam gadījuma lielumam, kad f_j ir varbūtības laikā a_j, $h(t)$ sevī ietver komponenti $h_j\sigma(t - a_j)$,kur

$$h_j = f + j/S(a_j).$$

Līdz ar to diskrētam sadalījumam ar elementiem f_j katrā laika punktā a_j, kur $a_1 < a_2 < ...$,

$$h(t) = \sum h_j\sigma(t - a_j),$$

kur

$$h_j = f_j/S(a_j) = f_j/(f_j + f_{j+1} + ...). \tag{1.2.6}$$

Nepārtraukti sadalītam gadījuma lielumam $h(t) = -S'(t)/S(t) = -d\ln S(t)/dt$ un tā kā $S(0) = 1$, tad

$$S(t) = \exp(-\int_0^t h(u)du) = \exp[-H(t)],$$

kur $H(.)$ sauc par integrēto risku. Ja $h(.)$ ir konstante ar vērtību ρ, sadalījums ir eksponenciāls,

$$S(t) = e^{-\rho t},\ f(t) = \rho e^{-\rho t}.$$

Diskrētam sadalījumam no (1.2.6) seko, ka

$$S(t) = \prod_{a_j<t}(1 - h_j).$$

Lai definētu integrētu risku diskrētam gadījumam, ērtuma labad veic logaritmisku transformāciju:

$$H(t) = \sum_{a_j<t}\ln(1 - h_j).$$

Līdz ar to izdzīvošanas funkcija ir izskatā: $S(t) = e^{-H(t)}$.

2. Neparametriskie novērtējumi

2.1. Stohastisks process izdzīvošanas analīzē

Īsi aprakstīsm galvenos jēdzienus, uz kā balstās izdzīvošanas procesu matemātiskā puse. Labu šīs teorijas izsklāstu sniedz Andersens [2]. Balstoties uz šī autora darbu, aprakstīsim šeit būtiskāko.

Aplūkojam stohastisku procesu $N(t), t \geq 0$, kuram izpildās īpašības:

1. $N(t) \geq 0$;
2. $N(t)$ ir skaitlis;
3. Ja $s \leq t$, atd $N(s) \leq N(t)$.

Ja $s \leq t$, tad $N(t) - N(s)$ ir notikumu skaits laika intervāla $(s, t]$. Tipisks piemērs ir Puasona process.

Fiksējam nepārtrauktu laika intervālu $\mathcal{T}$, kurš var būt formā $[0, \tau)$ vai $[0, \tau]$ kādam laika momentam τ, $0 < \tau \leq \infty$. $(\Omega, \mathcal{F})$ - mērojama telpa ar filtrāciju $(\mathcal{F}_t, t \in \mathcal{T})$, kas apmierina parastos nosacījumus (t.i. filtrācija ir augoša, pārtraukta no labās puses un pilna) katrai kopai $\mathcal{P}$ no varbūtību mēra. Dots stohastisks process $N = ((N_1, ..., N_k(t));\ t \in \mathcal{T}$, kurš definēts uz telpu $(\Omega, \mathcal{F})$ un adaptēts uz filtrāciju. Process apmierina multiplikatīvo jaudas procesu

$$\lambda(t) = h(t)Y(t),$$

kur $h(t)$ ir nenegatīva, deteministiska funkcija atkarīga no $P \in \mathcal{P}$ un $Y(t)$ - paredzams process kurš ir neatkarīgs no P. Bieži $h(t)$ ir funkcija, kas raksturo izmaiņu lielumu laikā, bet $Y(t)$ raksturo gadījumu skaitu, kas pakļauti riskam. Funkcija $h(t)$ mūsu gadījumā ir riska (hazard) funkcija un to var pierakstīt, izmantojot diferenciāli

$$h(t)dt = P(t \leq T < t + dt,\ C = 1 | T \geq t),$$

kur C ir cenzētu datu identifikators. Ja interesējošo izdzīvošanas laiku (nomirst pētāmā faktora, piemēram, slimības, ietekmē) apzīmē ar gadījuma lielumu X, bet ar Z izdzīvošanas laiku, kad novērojums ir pārtraukts vai nāve iestājusies ar pētāmo faktoru nesaistītos apstākļos, tad īstais novērojuma laiks ir $T = \min(X, Z)$. $C = 1$, ja $T = X$ un $C = 0$, ja $T = Z$.

2.2. Produkta integrālis

Aplūkosim produkta integrāļa jēdzienu tā, kā to apraksta [3] vai [2], kas tiek bāzēts uz Džila un Johansena (Gill and Johansen,1990) pētījumiem. Nodaļā par riska funkciju jau tika prezentēta formula, ar ko parasti saista produkta integrāļu rēķinus šajā jomā. Tā ir izdzīvošanas funkcija uzdota ar riska funkciju $S(t) = \prod(t_j < t)(1 - h_j)$. Vēl produkta integrālis tiek izmantots pie vislielākās ticamības funkcijas aprēķiniem.
Statistiskie modeļi bieži tiek balstīti uz to, ka mums ir nepārtraukts laika mainīgais, taču bieži prakstiski tas ir diskrēts lielums. Produkta integrāļa divi galvenie uzdevumi ir izveidot diskrēta laika reizinājumu viena soļa Markova pārvietošanās matricai un dot nepārtraukta laika risinājumu Kolmogorova diferenču vienādojumiem. Apskatīsim vispārīgu gadījumu.

Definīcija 1. [2] Dots $X(t),\ t \in \mathcal{T}$ ir $p \times p$ matrica ar kadlag(nepārtraukta no labās puses un ierobežota no kreisās) funkcijām no lokāli ierobežotām dispersijām. Definējam

$$Y = \prod(I + dX),$$

kas ir produktu integrālis no X pa intervāliem, kuri uzdoti formā $[0, t],\ t \in \mathcal{T}$ kā sekojoša $p \times p$ funkcija

$$Y(t) = \prod_{s \in [0,t]} (I + X(ds)) = \lim_{max|t_i - y_{[i-1]}| \longrightarrow 0} \prod (I + X(t_i) - X(t_{i-1})), \tag{2.2.1}$$

kur $0 = t_0 < t_1 < ... < T_n = t$ ir laiki no intervāla $[0, t]$ un matricas reizinājums tiek ņemts no kreisās puses uz labo. Kreisās galējās reizinājuma vērtības $X(0)$ jāaizvieto ar $X(0-) = 0$, jo galapunkts 0 arī ir iekļauts intervālā $[0, t]$.

Tas ir produkta integrālis tiek definēts $\mathcal{T}$ apakšintervālos. Pirmais reizinājuma elements ir intervāla robeža, kas vienmēr eksistē.
Kad X ir kāpņveida funkcija, tad produkta integrālis ir galīgs reizinājums pa X laika

lecieniem identitātes matricā plus pašas funkcija X pārvietojums - lecieni

$$Y = \prod (U + \Delta X).$$

Skalārā gadījumā, t.i., kad $p = 1$, reizinājuma kārtība nav svarīga un ar vieninieku var atdalīt lecienus starp X un to nepārtraukto daļu

$$\prod (1 + dX) = \exp(X^c) \prod (1 + \Delta X),$$

kur $X^c = X - \sum \Delta X$ un $\Delta X = X - X-$. Līdz ar to, ja X nav tikai skalārs, bet arī nepārtraukts, tad produkta integrālis ir eksponence $\prod (1 + dX) = \exp(X)$.
Svarīgākā produkta integrāļa īpašība ir spēja sareizināt produkta integrāļa elementus no savā starpā atdalītiem intervāliem, kas seko no definīcijas: katram $0 \leq s \leq t \leq u$ tiek atrasts

$$\prod_{(s,u)} (I + dX) = \prod_{(s,t]} (I + dX) \prod_{(t,u]} (I + dX).$$

Ne vien produkta integrālis eksistē, bet tas ir arī vienīgais atrisinājums noteiktā integrāļa vienādojumam.

Teorēma 1. *[2]* $\prod (I+dX)$ *eksistē un ir kadlag funkcija no lokāli ierobežotām dispersijām. Tas ir vienīgais atrisinājums integrālvienādojumam*

$$Y(t) = I + \int_{s \in [0,t]} Y(s-) X(ds).$$

Teorēma 2. *[2] Dots - S izdzīvošanas funkcija pozitīvam gadījuma mainīgajam T, t.i., $S(t) = P(T > t)$ visiem $t \geq 0$ un $S(0) = 0$. Definē kumulatīvo integrālo riska funkciju.*

$$H(t) = - \int \frac{S(ds)}{S(s-)}.$$

Tad

$$S(t) = \prod_{[0,t]} (1 - dA),$$

visiem t tādiem, ka $H(t) \leq \infty$.

Pierādījums. [2] Visiem t tādiem, ka $S(t-) > 0$ un no tā, ka $H(t) < \infty$ seko

$$S(t) = 1 - \int_0^t S(s-) H(s).$$

Izmantojot 1 definīciju, tagad $X = -H$ un $\mathcal{T} = t : S(t-) > 0$. Ja $\mathcal{T} = [0, \tau]$ kādam $\tau < \infty$, tad $S(\tau-) > 0$, bet $S(\tau) = 0$. Analoģiski, ja $H(t) = H(\tau) < \infty$ visiem $t \geq \tau$ un tas pats arī ar izdzīvošanas funkciju $S(t)$ visiem $t \geq \tau$. Ja tomēr $\mathcal{T} = [0, \tau)$ kādam $\tau \leq \infty$, tad varam parādīt, ka $H(t) = H(\tau) = \infty$ visiem $t \geq \tau$, un $S(t)$ nevar būt paplašināta un $t \geq \tau$. □

2.3. Nelsona-Alena modelis kā stohastisks process

Kā jau iepriekš ieviesām - mums ir stohastisks process $N = (N_1, ..., N_k)$ ar jaudas procesu $\lambda = (\lambda_1, ..., \lambda_k)$, kas apmierina multiplikatīvo jaudas modeli $\lambda(t) = h(t)Y(t)$. Lai atrastu novērtējumu

$$H(t) = \int_0^t h(s)ds, \tag{2.3.1}$$

tiek izmantots fakts, ka

$$M(t) = N(t) - \int_0^t h(s)Y(s)ds \tag{2.3.2}$$

ir lokāli kvadrātiski integrējams martingālis. Pārrakstot (2.3.2) diferenciāļos

$$dN(t) = h(t)dt + dM(t), \tag{2.3.3}$$

kur $dM(t)$ pieņemam kā troksni. Izteiksmi (2.3.3) ievietojot vienādojumā (2.3.1), iegūstam tā Nelsona-Alena novērtējumu

$$\hat{H}(t) = \int_0^t Y(s)^{-1}dN(s). \tag{2.3.4}$$

Ja $T_1 < T_2 < ...$ ir lēcienveida laika parametrs, kur pēc katra leciena izpildās notikums, tad (2.3.4) var parrakstīt kā summu

$$\hat{H}(t) = \sum_{j:T_j \leq t} Y(T_j)^{-1}. \tag{2.3.5}$$

$\hat{H}(t)$ ir augoša, nepārtraukta no labās puses kāpņveida funkcija ar pieaugumu $1/Y(T)$ katrā lēcienā T_j.

Statistikā Nelsona-Alena novērtējums izskatās formā

$$H(t) = \int_0^t h(s)J(s)ds, \tag{2.3.6}$$

kur $J(t) = I(Y(t) > 0)$. Izteiksme ir līdzīga (2.3.1) tikai tad, ja varbūtība, ka $Y(s) = 0,\ s \leq t$ ir maza. (2.3.6) var pārrakstīt kā dalījumu

$$H(t) = \int_0^t \frac{J(s)}{Y(s)} dN(s). \tag{2.3.7}$$

Piemērojot to izdzīvošanas datiem, kas ir cenzēti no labās puses, modeli varam uzrakstīt praktiskākos apzīmējumos. Pieņemsim, ka $X_1, ..., X_n$ ir neatkarīgi un vienādi sadalīti (iid) nenegatīvs gadījuma lielums ar absolūti nepārtārtrauktu sadalījuma funkciju F, riska funkciju $h = F_/(1 - F)$ un integrālo riska funkciju $H(t) = \int_0^t h(s)ds$. Mūs interesē dati, kas ir cenzēti no labās puses $(\tilde{X}_i, D), ß = 1, ..., n$, kur $\tilde{X}_i = X_i \wedge U_i$ un $D_i = I(\tilde{X}_i = X_i)$ kādam cenzētam laikam $U_1, ..., U_n$. Tad $N(t) = \sum_{i=1}^n I(\tilde{X}_i \leq t, D_i = 1)$ ir stohastisks process ar jaudas procesu $\lambda(t) = h(t)Y(t)$ ar $Y(t) = \sum_{i=1}^n I(\tilde{X}_i \geq t)$, ja dati ir neatkarīgi cenzēti no labās puses. Līdz ar to

$$J(t) = I(Y(t) > 0) = I(\tilde{X}_{(n)} \geq t),$$

kur $\tilde{X}_{(n)} = max(\tilde{X}_1, ..., \tilde{X}_n)$ un $J(t)/Y(t) \leq 1$ katram t.

2.4. Nelsona-Alena kodola funkcijas novērtējums

Praktiski piemēri kodola funkcijas novērtēšanā tiek aprakstīti Cleves [4] darbā par programmatūras Stata pielietojumu izdzīvošanas statistikā. Ar riska funkcijas gludināšanu ir strādājis Wang J.-L. Neparametriskās apstrādes metožu praktiskiem piemēriem palīdz viņa raksti [5] un [6]. Ilustratīvu ieskatu praktiskos aprēķinos var gūt Singera (Singer) un Villeta (Willett) darbā [7]. Gludināšanas piemērus apskata Bovmans(Bowman) un Acalini (Azzalini) [8].

Kodola funkcijas novērtējums riskam $h(t)$ ir balstīts uz $H(t)$ pieaugumu nogludināšanu. Definēsim šo novērtējumu

$$h(t) = b^{-1} \int_F K\left(\frac{t-s}{b}\right) d\hat{H}(s), \tag{2.4.1}$$

kur b ir joslas platums.

Šeit dotā kodola funkcija ir ierobežota intervālā $[-1,1]$ un tās integrālis ir vienāds ar 1. Ja procesam N uzdoti laika momenti/lecieni $T_1 < T_2 < ...$, tad (2.4.2) var pārrakstīt kā summu

$$h(t) = b^{-1} \sum_j K \left(\frac{t - T_j}{b} (Y(T_j) \right)^{-1}, \tag{2.4.2}$$

Funkcija tiek aprēķināta pa indeksiem j, tādiem, ka $t - b \leq T_j \leq t + b$. Atzīmēsim, ka $h(t)$ ir Nelsona-Alena svarots vidējais novērtējums no pieaugumiem $(Y(T_j))^{-1}$ intervālā $[t - b, t + b]$.

Daži kodola funkcijas piemēri:

1. Boxcar kodols $K(x) = 1/2 I_{|x| \geq 1}$;
2. Trīsstūra kodols $K(x) = (1 - |u|) I_{|x| \geq 1}$;
3. Gausa kodols $K(x) = \frac{1}{2\pi} e^{-\frac{x^2}{2}}$
4. Epaņečnikova kodols $K(x) = 0.75(1 - x^2)$.

2.5. Kaplana - Meijera kodola novērtējums

Vands (Wand) un Džons (Jones) [9] prezentēja arī Kaplana-Meijera kodola novērtējumu. Kodola blīvuma novērtējums f tiek izteikts caur Kaplana-Meijera empīrisko novērtējumu F^{KM}. Dots $X_1, ..X_n$ - necenzētie dzīves ilgumi, $Z_1, ..Z_n$ - cenzētie dzīves ilgumi. Abi mainīgie ir savā starpā neatkarīgi. Definēsim $Y_i = \min(X_i, Z_i)$ un $I = 1_{(X_i \leq Z_i)}$, $i = 1, ..., n$. Uzdevums ir novērtēt $f_X(x)$ - $X = (X_1, ..., X_n)$ blīvuma funkciju. Kaplana-Meijera novērtējumu var uzrakstīt kā sistēmu:

$$\hat{F}_X^{KM} = \begin{cases} 0, \ 0 \leq x \leq Y_{(j)} \\ 1 - \prod_{i=1}^{j-1} (\frac{n-i}{n-i+1})^{I_{(i)}}, \ Y_{(j-1)} < x \leq Y_{(j)}, \ j = 2, ..., n \\ 1, \ x > Y_{(n)} \end{cases},$$

kur $(Y_{(i)}, I_{(i)})$, $i = 1, ...n$ ir (Y_i, I_i) sakārtots augošā secībā pēc Y_i un ir F_X neparametrisks vislielākās ticamības novērtējums, kad dati ir cenzēti no labās puses. $f_X(x)$ kodola novērtējums izskatās formā

$$\hat{f}_X(x; b) = \int K_b(x - y) d\hat{F}_X^{KM}(y) = \sum_{i=1}^{n} s_i K_b(x - Y_{(i)}), \tag{2.5.1}$$

kur s_i ir $\hat{F}_X^{KM}$ novērtējuma leciena lielums $Y_{(i)}$ intervālā un b joslas platums.

2.6. Ticamības intervālu novērtējumi

Kaplana-Meijera ticamības intervālu novērtējums tika apskatīts darba sākumā. Tagad novērtēsim Nelsona-Alena ticamības intervālus. Atzīmēsim, ka $\hat{H}(t) = \hat{H}^{(n)}(t)$ ir vektors. Klasiska forma ir

$$\hat{H}(t) \pm c_{\alpha/2}\hat{\sigma}(t), \tag{2.6.1}$$

kur $c_{\alpha/2}$ ir standartizēta normāla sadalījuma kritiskā vērtība. Kā norāda literatūra [2], pec Monte Karlo simulācijām ir secināts, ka šis intervāla novērtējums ir neprecīzs maziem datu apjomiem. Labāku novērtējumu arī mazākām izlasēm var atrast ar delta metodi, kas jau tika apskatīta pie Kaplana-Meijera ticamības novērtējuma un 2.pielikumā ir pašas metodes izklāsts. Tātad, ja g ir kāda $H(t)$ tuva funkcija un $g'(x)$ ir nepārtraukta un atškirīga no nulles, kad $x = H(t)$, tad

$$\frac{g(\hat{H}(t)) - g(H(t))}{|g'(\hat{H}(s))|\hat{\sigma}(t)} \longrightarrow^{D} N(0,1),\ kad\ n \longrightarrow \infty.$$

No tā seko, ka asimptotiskais $100(1-\alpha)\%$ ticamības intervāls $g(H(t))$ funkcijai ir

$$g(\hat{H}(t)) \pm c_{\alpha/2}|g'(\hat{H}(s))|\hat{\sigma}(t), \tag{2.6.2}$$

kur

$$\hat{\sigma}^2_{ij}(t) = \sum_{t_i \leq t} \frac{dN_{ij}(t_k)}{Y_i(t_k)^2},\ k = 1, ..., n \tag{2.6.3}$$

Daudz pētījumu g izvēlē nav bijis, bet tiek ieteikti divi varianti -

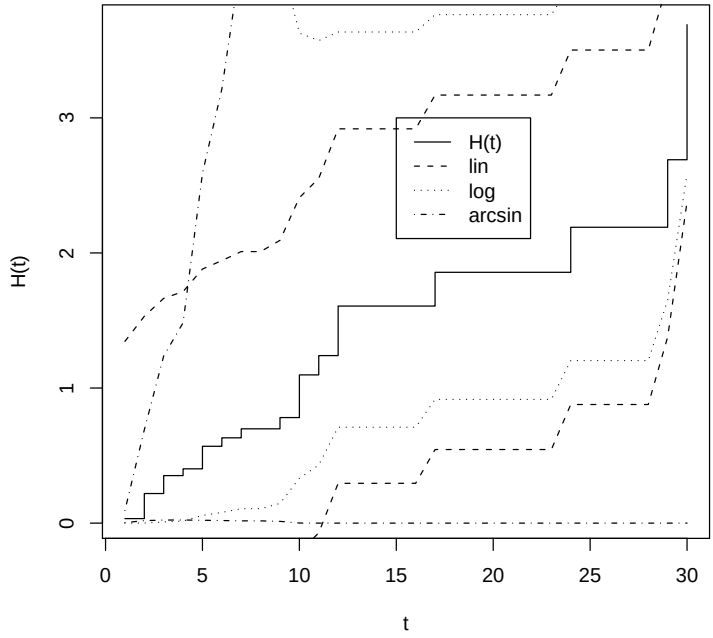
$$g(x) = \log x \tag{2.6.4}$$

$$g(x) = arcsin(\exp(-x/2)). \tag{2.6.5}$$

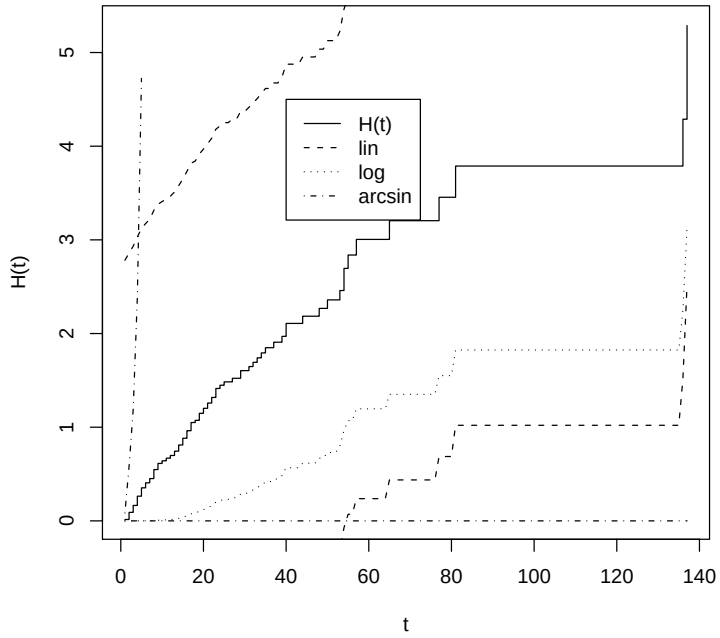
Šādi novērtējumi der pat izlasēm ar 25 elementiem un 50% cenzētiem datiem.

Vienlaicīgie ticamības intervāli. Visparīgi šeit jārunā par funkciju gludināšanu. Piemēram, plašu ieskatu var gūt no Randala(Randall)un Eubanka (Eubank) [10] vai Fans (Fan) un Gijbela (Gijbels) [11] un vēl daudzi citi. Andersen [2] prezentē Doksuma (Doksum) un Jendela (Yandell) ideju. Izmanto lielo izlašu īpašības. Izlaidīsim pašu izvirzījumu un pieņemsim galarezultātu:

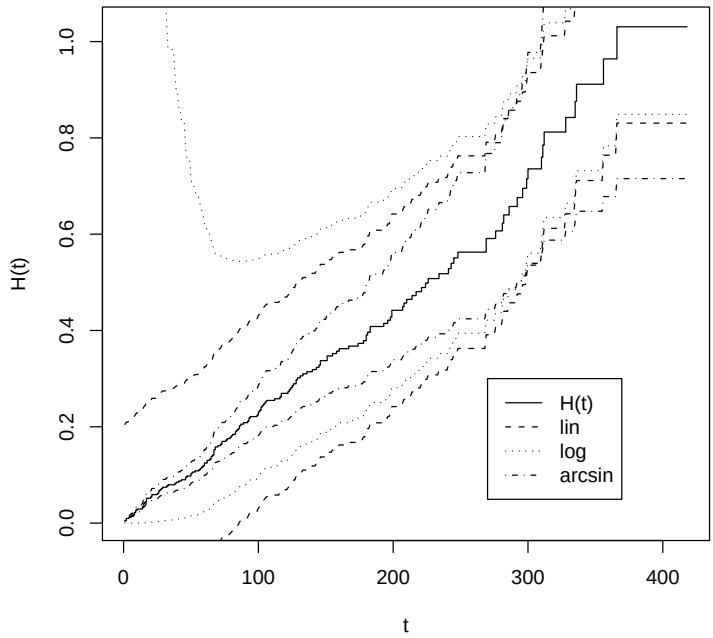
$$\hat{H}^{(n)}(t) \pm a^{-1}K_{q,a}(c_1, c_2)(1 - a^2\hat{\sigma}^2(s))/q\left(\frac{a^2\hat{\sigma}^2(s)}{1 - a^2\hat{\sigma}^2(s)}\right), \tag{2.6.6}$$



2.1. att.: Nelsona-Alena ticamības intervālu novērtējums *veteran* datiem, pirmie 30 elementi



2.2. att.: Nelsona-Alena ticamības intervālu novērtējums *veteran* datiem, pirmie 30 elementi



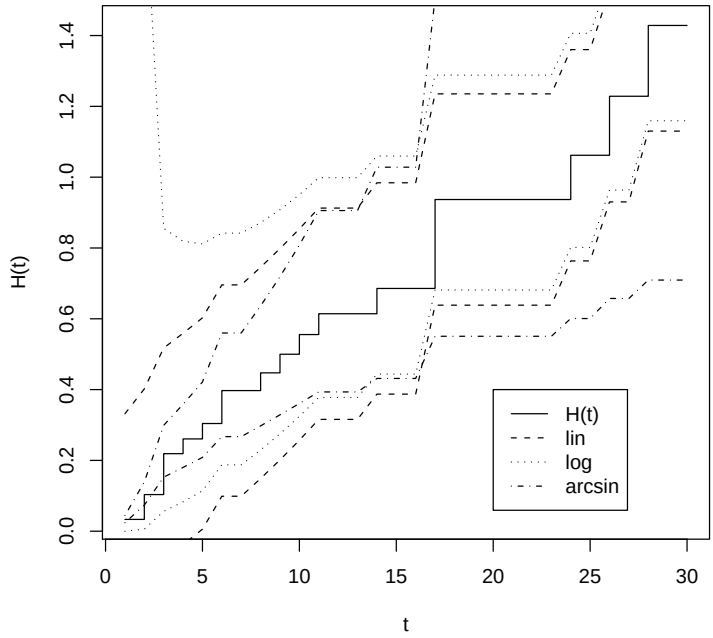
2.3. att. Nelsona-Alena ticamības intervālu novērtējums pbc datiem

kur a ir konstanšu vektors, kuru vērtības pieaug līdz ar vektora garumu, t.i. izlases lielumu. $K_{q,a}(c_1, c_2)$ ir kritiskā robeža sadalījumam.

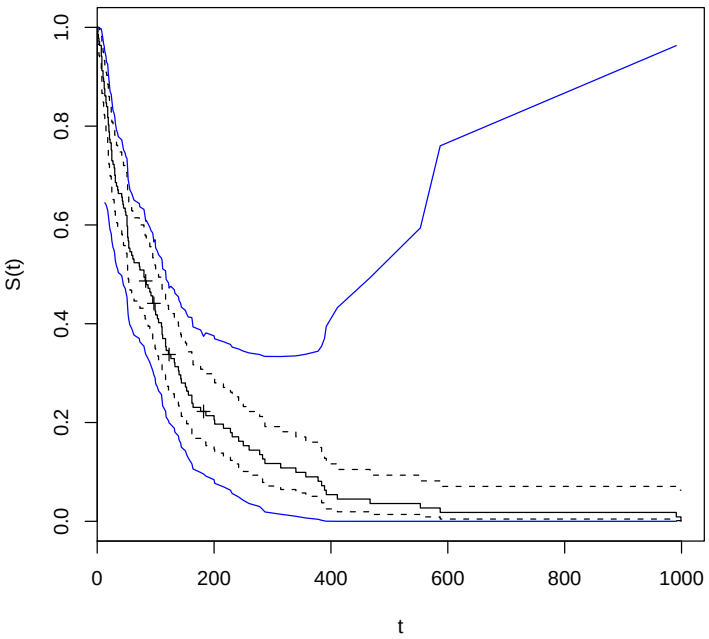
2.2. attēla (2.6.2) Nelsona-Alena punktveida ticamības intervāli ar log (2.6.4) un arcsin (2.6.5) transformāciju. Platākās joslas ir klasiskajam lineārajam novērtējumam (2.6.1). Pārbaudām, kā tas izskatās pie maziem apjomiem. Paņemam pirmos 30 elementus no *veteran* datiem. 2.1. attēlā redzami novērtējumi šai jaunajai izlasei. Aprēķinu algoritms pielikumā.

Atkārto novērtējumu aprēķinu *pbc* datiem(2.3. attēls). Novērojuma laika intervāla sākumā novērtējums ir visprecīzākais, jo risks nomirt vairumam pacientu vēl ir mazs. Ja paņem mazu izlasi (2.4. attēls)no šiem datiem, tad riska novērtējums ir precīzāks nekā *veteran* datu gadījumā.

Grūtāk atrast vienlaicīgos ticamības intervālus. Biežāk šim mērķim izmanto Hala un Velnera vienlaicīgos intervālus. To aparaksta, piemēram, [1] un [2]. Šo novērtējumu izmantojot Kaplana-Meijera ticamības intervālu konstruēšanai, rodas problēmas ar augšējo robežu - novērtējums nav pietiekami precīzs, to redzam 2.5. attēlā.



2.4. att. Nelsona-Alena ticamības intervālu novērtējums *pbc* datiem ar 30 elementiem



2.5. att.: Kaplana-Meijera novērtējums ar Hala un Velnera ticamības joslām veteran datiem

3. Izdzīvošanas funkciju salīdzināšana

Svarīgi ne vien novērtēt izdzīvošanas varbūtību, bet analizēt tās ietekmes faktorus. Medicīnā bieži tā tiek pētīta ārstēšanas metožu efektivitāte. Tāpat interesējošās grupas var būt dzimums, vecums un citi ar izdzīvošanu saistīti parametri. Vizuālu interpretāciju dažādu datu grupam attēlo ar Kaplana-Meiera novērtējumu palīdzību, kas ļauj gūt pirmo priekšstatu par datu atšķirību. Šo procdūru iespējams veikt vairumā programmatūru, arī R. 3.1. grafikā tiek attēlota mūsu datu *veteran* izdzīvošanas funkcijas atkarībā no laika divām grupām, kas raksturo dažādus ārstēšanas veidus vai pacienta personīgo ieradumus.

3.1. Log-ranga tests

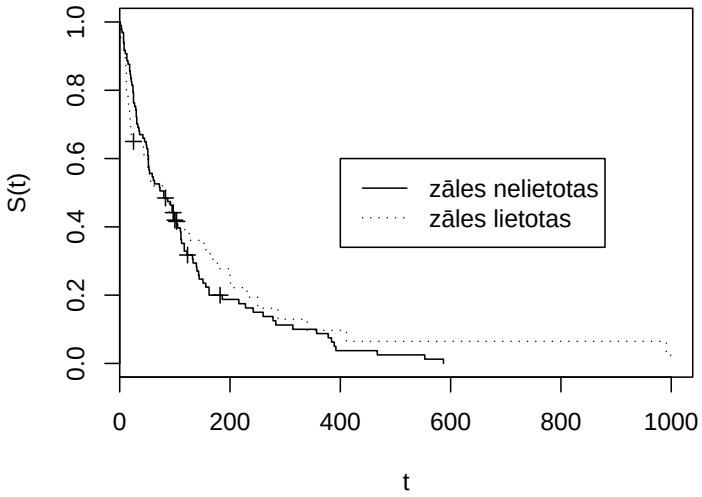
Divu izdzīvošanas līkņu salīdzināšanai visbiežāk izmanto statistisko hipotēžu testu, ko sauc par **log ranga testu**(log-rank test). To izmanto, lai testētu nulles hipotēzi, ka nav atšķirības starp pētāmā objekta izdzīvošanas līknēm (t.i., notikuma varbūtība katrā laika momentā abām grupām ir vienāda). Pieņemam, ka cenzora mainīgais nav atkarīgs no grupas. Testā tiek izmantots sagaidāmo nāves gadījumu skaits katrā grupā kādā laika momentā. Grupai 1 tas izsakāms ar formulu

$$\hat{e}_{1i} = \frac{n_{1i} d_i}{n_i}, \tag{3.1.1}$$

Visbiežāk izmanto pieņēmumu, ka nāves gadījumi ir ar hiperģometrisku sadalījumu, līdz ar to dispersija tādam lielumam ir

$$\hat{\nu}_{1i} = \frac{n_{1i} n_{0i} d_i (n_i - d_i)}{n_{1i}^2 (n_1 - 1))}. \tag{3.1.2}$$

Testu definē kā attiecību:



3.1. att.: Kaplana-Meijera izdzīvošanas funkcijas novērtējums divām cilvēku grupām. Redzam, ka cilvēkiem, kas lietojuši zāles ir nedaudz lielākas iespējas uz garāku mūžu. Dati- veteran

3.1. tabula Divu grupu notikumi

Notikums/Grupa	1	0	Kopā
Nāve iestājās	$d_1 i$	$d_0 i$	d_i
Nāve neiestājās	$n_1 i - d_1 i$	$n_0 i - d_0 i$	$n_i - d_i$
Riska grupa	$n_1 i$	$n_0 i$	n_i

$$Q = \frac{[\sum_{i=1}^{m} w_i(d_{1i} - \hat{e}_{1i})]^2}{\sum_{i=1}^{m} w_i^2 \hat{\nu}_{1i}}, \tag{3.1.3}$$

kur w_i - izlases svari, m - grupu skaits. Visbiežāk izmanto gadījumu, kad svari $w = 1$.

3.2. Koksa regresijas modelis

Plašu ieskatu par risku regresijas analīzi, izmantojot proporcionālo risku sniedz, piemēram, Kuiglijs (O'Quigley) [12]. Praktisku piemērus petījuši Therneau un Grambsch [13].

Apskatījām Kaplana-Meijera novērtējumu pa datu grupām 3.1. attēlā. Pieņēmām, ka katra grupa ir neatkarīga viena no otras un katrai varējām uzzīmēt savu izdzīvošanas funkciju. Taču pastāv iespēja, ka faktori, kas veido grupas, ir savā starpā saistīti. Regresijas uzdevums ir novērtēt šo saistību un izveidot modeli, kas vienas pētāmā objekta grupas izdzīvošanas iespējas izsaka ar kādas citas grupas datiem. Šeit runāsim par semiparametriskiem regresijas modeļiem, kad parametri tieši nav uzdoti, bet ir atkarīgi no viena mainīgā - laika t.

Izdzīvošanas laika sadalījumu var raksturot ar blīvuma funkciju, ja dots parametrisks sadalījums vai ar riska funkciju, ja par sadalījumu nekas nav zināms. Izsakot regresijas modeli ar riska funkciju, jānovērtē izteiksme

$$h(t, x, \beta) = h_0(t) r(x, \beta), \tag{3.2.1}$$

kur $h_0(t)$ parāda, kā izdzīvošanas funkcija mainās atkarībā no laika. Funkcija $r(x, t)$ raksturo izdzīvošanas funkcijas izmaiņas atkarībā no faktora mainīgā. Ja $h(t, x, \beta) = h_0(t)$, t.i. $r(x = 0, \beta) = 1$, tad $h_0(t)$ sauc par ir **bāzes riska funkciju**.

Divu parametru riska funkciju attiecība izsakāma

$$HR(t, x_1, x_0) = \frac{h(t, x_1, \beta)}{h(t, x_0, \beta)}.$$

No kā izriet, ka

$$HR(t, x_1, x_0) = \frac{h_0(t) r(x_1, \beta)}{h_0(t), r(x_0, \beta)} = \frac{r(x_1, \beta)}{r(x_0, \beta)}.$$

Deivids Koks(David Cox) bija pirmais, kas ieteica izmantot sakarību $r(x, \beta) = exp(x\beta)$. To zinot, riska funkcija pārrakstāma

$$h(t, x, \beta) = h_0(t) \exp(x, \beta) \tag{3.2.2}$$

un riska attiecība

$$HR(t, x_1, x_0) = \exp(\beta(x_1 - x_0)).$$

Šo sakarību sauc par **Koksa proporcionālo risku modeli** (Cox proportional hazards model).

Tagad ar to izteiksim izdzīvošanas funkciju:

$$S(t, x, \beta) = \exp(-H(t, x, \beta)),$$

kur $H(t, x, \beta)$ ir kumulatīva riska funkcija laikā t pētījuma objektam ar kovarianti x. Pieņēmām, ka izdzīvošanas laiks ir nepārtraukts, tāpēc kumulatīvo riska funkciju var pierakstīt kā integrāli:

$$H(t, x, \beta) = \int_0^t h(u, x, \beta)du = r(x, \beta) \int_0^t h_0(u)du = r(x, \beta)H_0(t). \tag{3.2.3}$$

Līdz ar to izdzīvošanas funkcija ir $S(t, x, \beta) = \exp(-r(x, \beta)H_0(t)$. Tā kā $H_0(t)$ ir bāzes kumulatīvā riska funkciju, tad ieviešam bāzes izdzīvošanas funkciju $S_0(t) = e^{-H_0(t)}$ un izdzīvošanas funkciju varam pārrakstīt

$$S(t, x, \beta) = [S_0(t)]e^{x\beta} \tag{3.2.4}$$

No šejienes redzam, ka izdzīvošanas funkcija parāda sakarību starp bāzes riska funkciju un eksponenciālu funkciju, kas apraksta kovariansu ietekmi. Interpretēsim formulu ar piemēru, par kovarianti izvēloties vecumu, ko apzīmēsim ar a. Tad $x = a - \bar{a}$. Bāzes idzīvošanas funkcija attiecas uz to pētāmo objektu, kura vecums $\bar{a}$ sakrīt ar vidējo vecumu visā izlasē. Pieņemot, ka risks izdzīvot, pieaugot vecumam samazinās, t.i. $\beta > 0$ un katrs $a < \bar{a}$, no tā seko $x > 0$, $\exp(x\beta) > 1$ un $S(t, x, \beta) < S_0(t, x, \beta)$. Tātad iespēja izdzīvot ir mazāka vecumā a nekā $\bar{a}$. Tas ir katrā laika momentā tā daļa novērojamo objektu, kas ir dzīvi vecumā a ir mazāka nekā vecumam $\bar{a}$. Un otrādi - ja pieņem, ka $a < \bar{a}$, tad $x < 0$, $\exp(x\beta) < 1$ un $S(t, x, \beta) > S_0(t, x, \beta)$.

Piemērs 1. Atradīsim proporcionālo risku mūsu datiem *veteran*. Tālāk tiks apskatīts β koeficienta atrašana, šobrīd pieņemsim, ka to zinām. R programma piedāvā iebūvētu funkciju, lai atrastu šos koeficientus. Noskaidrosim vecuma ietekmi uz izdzīvošanu. x_1 mums ir vecums 70 gadi, $x_0 = 58$ - vidējais vecums starp visiem izlases cilvēkiem. Atrodam, ka $\beta = 0.0075$, tad

$$HR = \exp(0.0075(70 - 58)) = 1.092,$$

kas norāda, ka līdz ar vecumu pieaug risks nomirt, bet atšķirība starp 60 gadiem un vidējo vecumu risku nav liela.

3.2.1. Koeficientu noteikšana

Mūsu modelis izskatās formā

$$y = \beta_0 + \beta_1 x + \sigma\epsilon \tag{3.2.5}$$

Lai atrastu β koeficientus, izmanto ticamības funkciju (likelihood function). Neapskatīsim šeit visu formulu veidošanās mehānismu, bet tikai galvenās detaļas. Ticamības funkciju var izteikt kā necenzēto datu blīvuma funkcijas reizinājumu ar cenzēto datu izdzīvošanas funkciju

$$[f(t, \beta, x)]^c \times [S(t, \beta, x)]^{1-c},$$

kur c ir mainīgais, kas raksturo datus - $c = 0$ - dati necenzēti, $c = 1$ - cenzēti. Ja mums ir n neatkarīgi novērojumi, tad ticamības funkcija ir (3.2.5) reizinājums:

$$l(\beta) = \prod_{i=1}^{n}([f(t_i, \beta, x_i)]_i^c \times [S(t_i, \beta, x_i)]^{1-c_i}) \tag{3.2.6}$$

No izteiksmes (1.2.5) izsakām

$$f(t, x, \beta) = h(t, x, \beta) \times S(t, x, \beta).$$

To ievietojot izteksmē (3.2.6) iegūstam, ka

$$l(\beta) = \prod_{i=1}^{n}([h(t_i, x_i, \beta) \times S(t_i, x_i, \beta)]^{c_i} \times [S(t_i, x_i, \beta)]^{1-c_i}),$$

ko vienkāršojot iegūst:

$$l(\beta) = \prod_{i=1}^{n}([h(t_i, x_i, \beta)]^{c_i} \times [S(t_i, x_i, \beta)]).$$

Lai noteiktu maksimālo ticamību attiecībā pret β, maksimizēsim šīs funkcijas log-ticamības funkciju. Vienlaicīgi riska un izdzīvošanas funkciju vietā ievietosim to funkcijas izteiktas ar bāzes funkcijām, no izteiksmēm (3.2.2) un (3.2.4)

$$L(\beta) = \sum_{i=1}^{n}(c_i \ln[h_0(t_i)] + c_i x_i \beta + e^{x_i\beta} \ln[S_0(t_i)])$$

Tā kā funkcija no logaritmiskas transformācijas ir monotona, tad gan log-ticamības funkcijas, gan (1.1.4) sasniegs savu maksimumu pie viena un tā paša koeficienta β. Log-ticamības funkcijas priekšrocība ir salīdzinoši vieglāka vienādojuma atrisināšana.

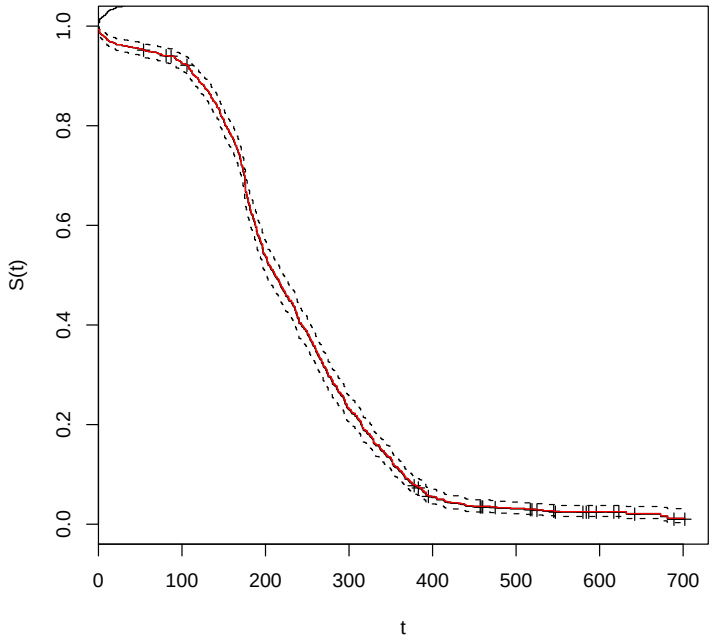
4. Latvijas tuberkulozes pacientu datu analīze

R programmā iebūvētos datus aizstāsim ar informāciju par Latvijas pacientiem un veiksim tās analīzi. Nepieciešami dati, kuri apraksta slimniekus ar nāvējošu slimību - attiecīgi starp datiem jābūt reģistrētiem nāves gadījumiem. Šī iemesla dēļ tika izvēlēta Tuberkulozes un plaušu slimību valsts aģentūra. Iegūti dati ir par 2007. gadā reģistrētajiem pacientiem. Datus veido parametri - slimības reģistrēšanas laiks, personas dzimšanas datums, dzimums, pacienta nodarbošanās, ārstēšanas rezultāts, ārstēšanas efektivitātes datums (brīdis, kad veic novērojumu par pacienta stāvokli - parasti šāda informācija datu bāze tiek reģistrēta pēc gada kopš uzskaites datuma), riski - alkohols, narkotikas, cietums. Pavisam ir 994 ieraksti ar derīgiem elementiem. Starp tiem ir 50 ieraksti jeb 5% ar cenzētiem datiem, 61 pacienti jeb 6% - miruši. Parējie ir izārstēti vai joprojām tiek ārstēti.

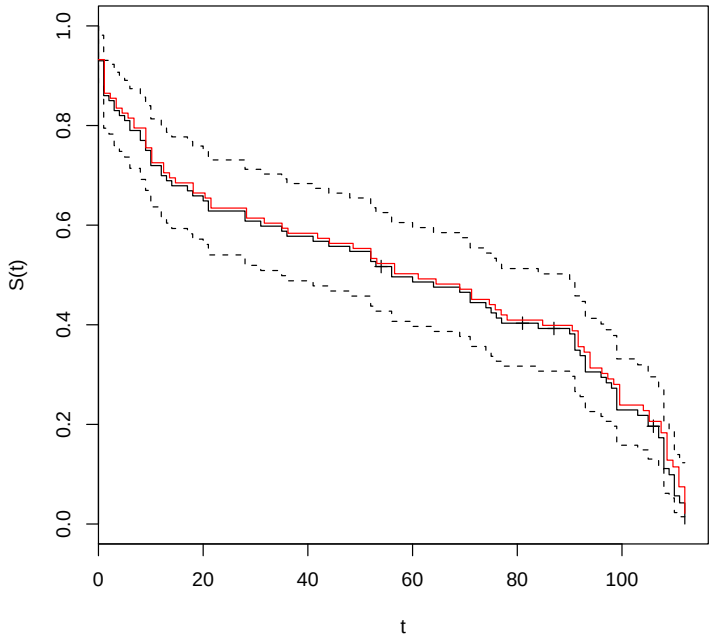
Sāksim ar Kaplana-Meijera novērtējumu. 4.1. attēlā vienā grafikā ar Kaplana-Meijera novērtējumu ir arī caur Nelsona-Alena izteiktā izdzīvošanas funkcija. Izlases apjoms ir liels, un atšķirība ir nepamanāma, tāpēc paņemsim pirmos 100 novērtējumus(4.2. attēls), lai secinātu, ka Nelsona-Alena novērtējums ir lielāks.
Novērtēsim riska funkciju - 4.5. attēlā. Ticamības intervālu novērtējums ir ļoti tuvs funkcijai, jo nāves gadījumu izlasē ir relatīvi maz, tāpat arī cenzētu ierakstu. 4.4. attēlā izdzīvošanas funkcijas pa ietekmes faktoriem.

Kopumā atšķirības starp grupām ir mazas. Izteiktāk var novērot narkotiku lietotāju lielāku mirstības risku. Taču jāņem vērā, ka to proporcija visā izlasē ir maza un tas var būt par šo atšķirību iemeslu.

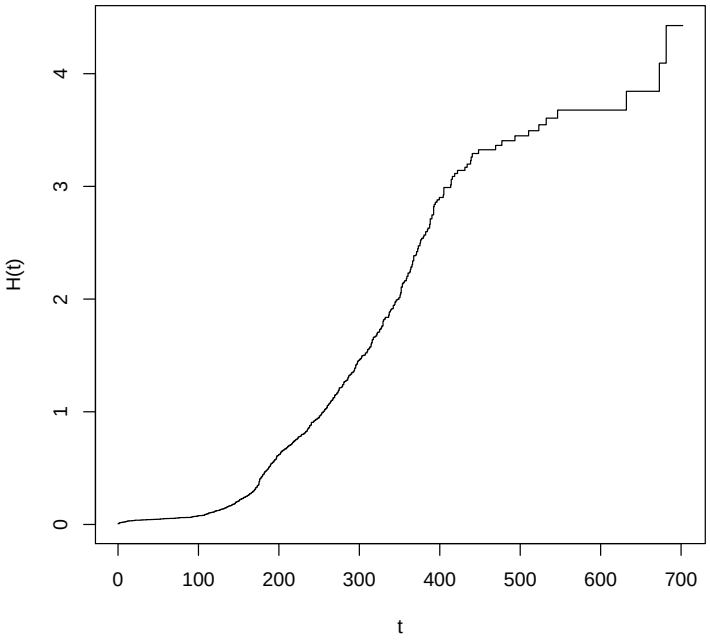
Analizējot uz kodola novērtējuma grafiku 4.5. attēlā jāatceras, ka pētām mirstību un visbiežāk slimība pie letālām sekām noved pēc kāda ilgāka perioda, tāpēc mazi novērtē-



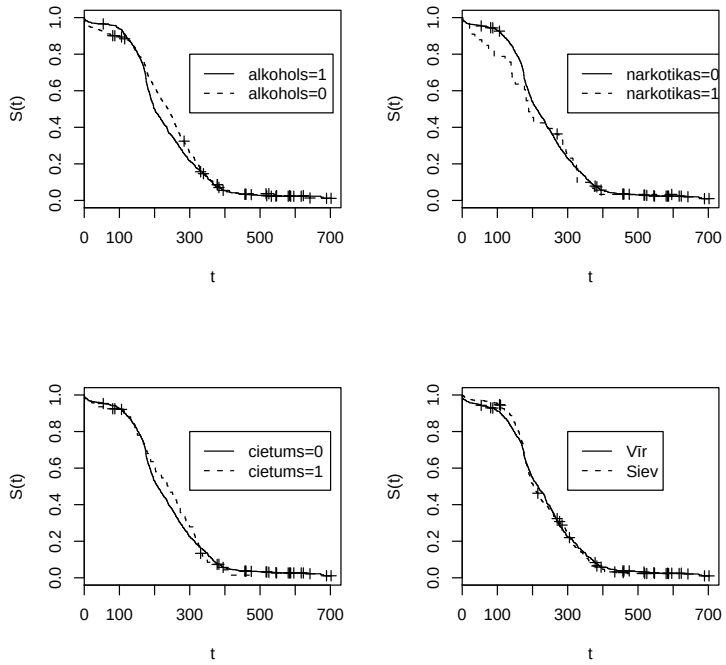
4.1. att.: Kaplana-Meijera un Nelsona-Alena izdzīvošanas funkcijas novērtējumi. Atšķirības ir maz redzamas.



4.2. att.: Kaplana-Meijera un Nelsona-Alena izdzīvošanas funkcijas novērtējumi. Izlase ar 100 elementiem - uzskatāmāk redzamas atšķirības, kuras šajā gadījumā ir ļoti mazas



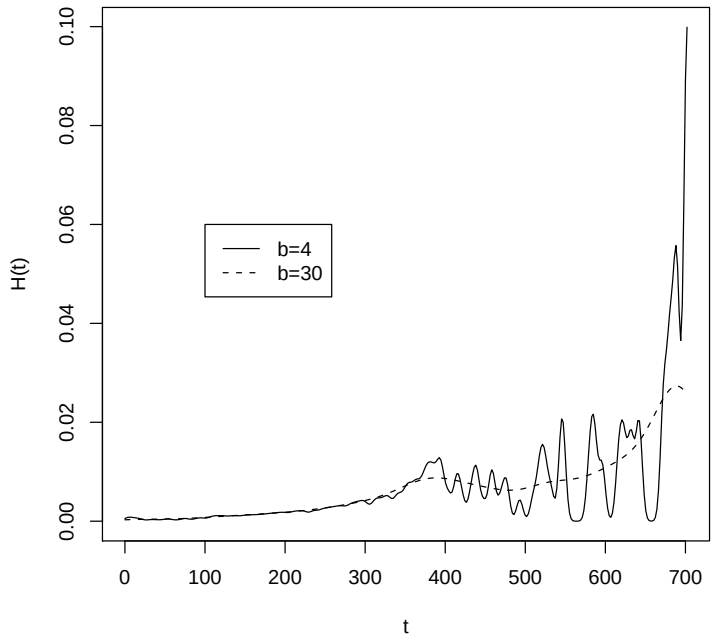
4.3. att. Nelsona-Alena novērtējums



4.4. att. Kaplana-Meijera novērtējumi salīdzinot pa grupām

4.1. tabula Iespējamo ietekmes faktoru proporcija izlasē

Faktors			
Dzimums	vīrietis	671	68 %
	sieviete	323	32%
Alkohols	lieto	308	30%
	nelieto	693	70%
Narkotikas	lieto	34	3 %
	nelieto	961	97 %
Cietumā	ir	93	9%
	nav	901	91 %



4.5. att.: Nelsona-Alena kodola novērtējums ar krosvalidācijā iegūto joslas platumu - 4 un par joslas platumu izvēloties vienu mēnesi - tas ir 30 diena

juma apgabali šeit neder. Bieži joslas platumam izvēlas mēnesi. Var novērot, ka lielāks risks nomirt ir pēc apmēram gada kopš konstatēta slimība. Kodola konstruēšanas R kods pielikumā.

Rezultāti un secinājumi

Pētot ar izdzīvošanas analīzi saistīto literatūru, var secināt, ka būtiska loma ir riska funkcijām un to novērtējumiem. Kā divu riska funkciju attiecība izsakāms Koksa proporcionālo risku modelis, ar ko savukārt var novērtēt vairumu citu statistiku. Koksa proporcionālo modeli vistiešāk saistīts ar izdzīvošanas statistikas regresiju analīzi.
Pētot grāmatas un publikācijas var secināt, ka daudzi autori atsaucas uz Andersena, Borgana, Gill un Keidinga darbu [2], kurā izdzīvošans statistikas modeļi pētīti no stohastisko procesu puses. Neparametrisko modeļu pētīšanā nozīmīgs ieguldījums ir Wanga [9], risku regresiju analīzē - O‘ Kuiglijs [12]. Senāka klasika ir Oaks (Oakes) un Koks (Cox)[3].
Līdzās riska novērtējumiem un izdzīvošanas funkcijai svarīgi izprast produktu integrāļu teoriju, jo uz tās balstās izdzīvošanas funkcija uzdota ar riska funkciju un vislielakās ticamības funkcijas aprēķini, kas nepieciešami regresiju modelēšanā. Parasti šajā tēmā neiztiek bez atsaucēm uz Gilu(Gill) un Johansenu(Johansen).

Tuberkulozes datu apstrāde parādīja, ka uzskatāmiem izdzīvošanas statistikas pielietojuma rezultātiem vairāk der dati, kuros ir lielāks mirstības biežuma rādītājs - no otras puses - tā ir statistika, kurā labāk samierināties ar mazāk vizuāli interesantiem rezlutātiem nekā konstatēt, ka izdzīvošanas varbūtībai kādai noteiktai petāmo objektu kopai ir maza. Šajā gadījumā ar to jāsaskaras - datu ir daudz, bet ticamības intervāliem grafiskai novērēšanai labāk ņemt daļu no tiem, lai redzētu kā novērtējumi un to ticamības intervāli mainās laika gaitā. Pretējā gadījumā joslas ir pārāk šauras un neuzskatāmas.

Nākamais solis būtu apskatīt riska funkciju gludināšanu, vienlaicīgo ticamības joslu kostruēšanu, iespējams, izmantojot Lehmana alternatīvo modeli. Neparametrisko metožu pielietojumam klāt pielikt butsrapa metodi Nelsona-Alena novērtējumu iegūšanai.

Pateicības

Vislielākā pateicība vecākiem par to vērtību sistēmu, kas vienmēr ir palīdzējusi saprast arī sarezģītu un grūtu situāciju jēgu.
Paldies darba vadītājam Jānim Valeinim, kas spēj atrast motivējošu pieeju darbam un cilvēkam. Un dr.Vijai Riekstiņai par atsaucību, palīdzot ar piemēram nepieciešamo datu iegūšanu.

Izmantotā literatūra un avoti

[1] Hosmer David W. and Lemeshow S. *Applied Survival Analysis.* John Wiley and Sons, Inc, New York, 1999.

[2] Gill Richard D. Keiding Niels Andersen Per Kragh, Borgan rnulf. *Statistical models based on counting processes.* Springer, New York, 1993.

[3] Cox D.R. and Oakes D. *Analysis of survival data.* Chapman and Hall, London, 1984.

[4] Gutierrez Roberto Cleves Mario Alberto, Gould William. *An Introduction to Survival Analysis Using Stata, Second Edition.* Stata Press, Texas, 2008.

[5] Wang J.-L. *Smoothing hazard rate. Encyclopedia of Biostatistics, 2nd Edition, Vol 7.* John Wiley and Sons, Inc, New York, 2005.

[6] Wang J.L. Mller, H.G. An invariance principle for discontinuity estimation in smooth hazard functions under random censoring. *Sankhya A*, 58:392–402, 1996.

[7] Willett John B. Singer, Judith D. *Applied longitudinal data analysis: modeling change and event occurrence.* Oxford University Press, US, 2003.

[8] Adelchi Azzalini A. W. Bowman. *Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrationsl.* Oxford University Press, US, 1997.

[9] M. Chris Jones Matt P. Wand. *Kernel smoothing.* Chapman and Hall/CRC, US, 1995.

[10] Randall L. Eubank. *Nonparametric regression and spline smoothing.* Chapman and Hall/CRC, US, 1999.

[11] Ir?ne Gijbels Jianqing Fan. *Local polynomial modelling and its applications.* CRC Press, US, 1996.

[12] O'Quigley J. *Proportional Hazards Regression*. Springer, New York, 2008.

[13] Patricia M. Grambsch Terry M. Therneau. *Modeling survival data: extending the Cox model*. Springer, New York, 2000.

5. Pielikums

5.1. Krtiskās vērtības Hala-Vernera ticamības intervālu aprēķinam

5.1. tabula Krtiskās vērtības Hala-Vernera ticamības intervālu aprēķinam

	$\hat{\alpha} = n\hat{\sigma}^2(t_m)/[1 + n\hat{\sigma}^2(t_m)]$							
1-α	0.1	0.25	0.40	0.50	0.60	0.75	0.90	1.0
0.90	0.599	0.894	1.062	1.133	1.181	1.217	1.224	1.224
0.95	0.682	1.014	1.198	1.273	1.321	1.354	1.358	1.358
0.99	0.851	1.256	1.470	1.552	1.600	1.626	1.628	1.628
	$\hat{\sigma}^2(t_m) = \sum_{t_i} \frac{d_i}{n_i(n_i - d_i)}$							

5.2. Delta metode

Apraksts balstās uz Hosmera [1] un Andersena [2] darbiem.

Problēmas pamatā - kā, izmantojot parametra novērtējumu, konstruēt dispersijas novērtējumu. Katram novērtējumam nepieciešams ticamības intervāls vai hipotēzes testi.

Delta metode jau izsenis ir pazīstama kā dispersijas novērtējuma atrašanas metode. Tās pamatideja ir Teilora rindas izvirzījums, ar kuras palīdzību tiek aproksimētas sarežģītas funkcijas.

Lai funkciju aproksimētu ar Teilora rindu, ir nepieciešama gluda funkcija. Lai redzētu delta metodes pielietojumu, izmantosim pirmos divus Teilora rindas locekļus mainīgā vidējā

aproksimācijai.

$$f(X) = f(\mu) + (X - \mu)f'(\mu), \tag{5.2.1}$$

kur $f\prime(\mu) = \frac{\delta f(X)}{\delta X}|_{X=\mu}$ ir funkcijas atvasinājums attiecībā pret X novērtēta pie vidējās vērtības no X. No (5.2.1) seko, ka funkcijas dispersija ir aptuveni

$$D[f(X)] \cong D(X - \mu)[f'(\mu)]^2 \cong \sigma^2[f'(\mu)]^2, \tag{5.2.2}$$

kur σ^2 ir dispersija no X. Delta metodes novērtējums definē funkcijas dispersiju gadījuma, kad izmanto vidējās vērtības un dispersijas novērtējumus $\hat{\mu}$ un $\hat{\sigma}^2$

$$\hat{D}[f(X)] \cong \hat{\sigma}^2[f'(\hat{\mu})]^2. \tag{5.2.3}$$

Piemēram, funkcija, kas tiek izmatota darbā - $\ln(X)$. (5.2.1) izvirzījums ir

$$\hat{D}[\ln(X)] \cong \hat{\sigma}^2 \frac{1}{\hat{\mu}^2}. \tag{5.2.4}$$

Kā otro piemēru, izvirzīsim Kaplana-Meijera izdzīvošanas funckijas log novērtērtējuma dispersiju. Izdzīvošanas funkcijas novērtējums ir

$$\hat{S}(t) = \prod_{t_{(i)} \leq t} \frac{n_i - d_i}{n_i}$$

un tā log novērtējums

$$\ln[\hat{S}(t)] = \sum_{t_{(i)} \leq t} \ln(\frac{n_i - d_i}{n_i}) = \sum_{t_{(i)} \leq t} \ln(\hat{p}_i),$$

kur $\hat{p}_i = (n_i - d_i)/n_i$. Pirmais pieņēmums, dispersijas novērtējuma konstruēšanā, ir tas, ka novērojumi n_i ir savā starpā Bernulli neatkarīgi ar konstantu varbūtību p_i. Konstantās varbūtības novērtējums ir $\hat{p}_i$ ar dispersijas novērtējumu $\hat{p}_i(1 - \hat{p}_i)/n_i$. Teilora rindas izvirzījums log funkcijai (5.2.4) dod

$$\ln(\hat{p}_i) \cong \ln(p_i) + (\hat{p}_i - p_i)\frac{1}{\hat{p}_i}$$

un no (5.2.3) delta metodes dispersijas novērtējums ir

$$\hat{D}[\ln(\hat{p}_i)] \cong \frac{\hat{p}_i(1 - \hat{p}_i)}{\hat{p}_i^2 n_i} \cong \frac{d_i}{n_i(n_i - d_i)}.$$

Otrais pieņēmums ir tas, ka novērojumi dažādās riska grupās ir neatkarīgi. Līdz ar to delta metodes Kaplana-Meijera log novērtējuma dispersijas novērtējums ir

$$\hat{D}(\ln[\hat{S}(t)]) \cong \sum_{t_{(i)} \leq t} \hat{D}[\ln(\hat{p}_i)] \cong \sum_{t_{(i)} \leq t} \frac{d_i}{n_i(n_i - d_i)}.$$

Tālāk, lai atrastu Kaplana-Meijera novērtējumu jāatceras,ka

$$f(X) = \exp(X),$$

t.i. $\hat{S}(t) = \exp(\ln[\hat{S}(t)])$. Rindas izvirzījums attiecīgi

$$\exp(X) \cong \exp(\mu) + (X - \mu)\exp(\mu)$$

un no (5.2.2) dispersijas tuvinājums ir

$$\hat{D}[\exp(X)] \cong \sigma^2[\exp(\mu)]^2. \tag{5.2.5}$$

Tuvinājuma (5.2.5) pielietojums ir Grīnvuda(Greenwood) novērtējums

$$\hat{D}[\hat{S}(t)] \cong [\hat{S}(t)]^2 \sum_{t_{(i)} \leq t} \frac{d_i}{n_i(n_i - d_i)}.$$

Kaplana-Meijera ticamības intervālu novērtējums balstās uz log-log izdzīvošanas funkciju, t.i., $\ln(-\ln[\hat{S}(t)])$. Šī novērtējuma dispersijas novērtējums ir

$$\hat{D}(\ln[-\ln(\hat{S}(t))] \cong \frac{1}{[\ln(\hat{S})(t))]^2} \sum_{t_{(i)} \leq t} \frac{d_i}{n_i(n_i - d_i)}$$

.

5.3. Izveidoto programmu kods

```
library(splines)\\
library(survival)
library(crq)
library(Zelig)
library(timereg)
library(muhaz)
--------------------------- Dati ---------------------------
attach(veteran)
data<-veteran
drug<-data[,1]
status<-data[,4]
age<-data[,7]
```

```
time<-data[,3]
____________________________Reprezentācija ______________________
____________________________Kaplan Meier novērtējums _______
m.dat<-matrix(c(time,status,age,drug),length(data[,1]),4)
sort<-m.dat[order(m.dat[,1]), ]
ft<-ftable(time,status)
d<-ft[,2]
c<-ft[,1]
p<-c()
s<-c()
n<-c()
var_p<-c()
var_s<-c()
cu<-c()
cl<-c()
z<-1.96
n[1]<-length(data[,1])
var_p[1]<-d[1]/(n[1]*(n[1]-d[1]))
p[1]<-(n[1]-d[1])/n[1]
s[1]<-p[1]
var_s[1]<-var_p[1]/(log(s[1]))^2
cu[1]<-log(-log(s[1]))+z*sqrt(var_s[1])
cl[1]<-log(-log(s[1]))-z*sqrt(var_s[1])
for(j in 1:(length(d)-2)){
n[j+1]<-n[j]-d[j]-c[j]
s[j+1]<-p[j]*s[j]
p[j+1]<-(n[j+1]-d[j+1])/n[j+1]
var_p[j+1]<-d[j+1]/(n[j+1]*(n[j+1]-d[j+1]))+var_p[j]
var_s[j+1]<-var_p[j+1]/(log(s[j+1]))^2
cu[j+1]<-log(-log(s[j+1]))+z*sqrt(var_s[j+1])
cl[j+1]<-log(-log(s[j+1]))-z*sqrt(var_s[j+1])
j<-j+1
```

```
}
#tā kā pēdējais ir '1' un tas nomirst
s[j+1]<-0
plot(unique(sort[,1]),s[1:length(s)],type="s",xlab="t",ylab="S(t)")
lines(unique(sort[,1]),s[1:length(s)],type="s",xlab="t",ylab="S(t)",col="blue")
dev.print(device=postscript, "graf1.eps", onefile=FALSE,horizontal=FALSE)
_________________________  Ticamības intervāli KM novērtējumam
#Pointwise conf int
#lines(unique(sort[,1]),exp(-exp(cl[1:length(cl)])),col='green')
#lines(unique(sort[,1]),exp(-exp(cu[1:length(cu)])),col='green')
rind<-unique(sort[1:length(sort[,1])-1,1])
lines(rind,exp(-exp(cl[1:length(cl)])),col='green')
lines(rind,exp(-exp(cu[1:length(cu)])),col='green')
##to pašu iebūvēti
plot(survfit(Surv(time,status),data=data),ylab="S(t)",xlab="t")
dev.print(device=postscript, "graf2.eps", onefile=FALSE,horizontal=FALSE)
#Hall and Wellner bands
nn<-length(data[,1])
a<-nn*var_p[j]/(1+nn*var_p[j])
# no 1.pielikuma tabulas, pie a=0.98, H=1.358
H<-1.358
bu<-c()
bl<-c()
length(s)
for(j in 1:length(d)){
bu[j]<-log(-log(s[j]))+(H*(1+nn*var_p[j])/(sqrt(nn)*abs(log(s[j]))))
bl[j]<-log(-log(s[j]))-(H*(1+nn*var_p[j])/(sqrt(nn)*abs(log(s[j]))))
j<-j+1
}
lines(unique(sort[,1]),exp(-exp(bl[1:length(bl)])),col='blue')
lines(unique(sort[,1]),exp(-exp(bu[1:length(bu)])),col='blue')
________________________  Nelson Aalen novērtējums _____________________
```

```
p<-c()
s<-c()
n<-c()
h<-c()
H<-c()
n[1]<-length(data[,1])
p[1]<-(n[1]-d[1])/n[1]
s[1]<-p[1]
H[1]<-d[1]/n[1]
h[1]<-d[1]/length(data[,1])
for(j in 1:(length(d)-1)){
n[j+1]<-n[j]-d[j]
s[j+1]<-p[j]*s[j]
p[j+1]<-(n[j+1]-d[j+1])/n[j+1]
h[j+1]<-d[j+1]/length(data[,1])
H[j+1]<-(d[j]/n[j])+H[j]
j<-j+1
}
par(mfrow=c(2,1))
plot(sort(age),exp(-H[1:length(H)]),type="s",xlab="vecums",ylab="NA riska novērtēju
plot(sort(time),exp(-H[1:length(H)]),type="s",xlab="laiks",ylab="NA riska novērtēju
legend(400,0.6,c("Kaplana-Meijera","Nelsona-Alena"),lty=c(1,1),col=c("black","red")
dev.print(device=postscript, "graf.eps", onefile=FALSE,horizontal=FALSE)

_________________________  Nelson Aalen novērtējuma kodolgludināsana _______________
y<-time
fit.surv <- survfit(Surv(y,status))
xx<-seq(min(y),max(y),len=length(y))
A <- function(time.point){
sum(fit.surv$n.event[fit.surv$time <= time.point]/
fit.surv$n.risk[fit.surv$time <= time.point])
}
```

```
for(i in 1:(length(y))){
al[i]<-A(xx[i])
}
bb<-hcv(y)
b<-bb
m<-(max(y)-min(y))/b
n<-fit.surv$n.risk
r<-fit.surv$n.event
tt<-seq(min(y),max(y),len=length(r))
T<-fit.surv$time
kod<-function(t)
{
1/b*sum(dnorm((t-T)/b)*n^(-1))
}
novk<-c()
for(i in 1:(length(r))){
novk[i]<-kod(tt[i])
}
plot(tt,novk,type="l",xlab="t",ylab="H(t)")
```

Maģistra darbs " Izdzīvošanas analīze ar pielietojumu medicīniskajā statistikā" izstrādāts LU Fizikas un Matemātikas fakultātē.

Ar savu parakstu apliecinu, ka pētījums veikts patstāvīgi, izmantoti tikai tajā norādītie informācijas avoti un iesniegtā darba elektroniskā kopija atbilst izdrukai.

Autors: Olga Grakoviča

_____ _____

(paraksts) (datums)

Rekomendēju darbu aizstāvēšanai.

Vadītājs: doc. Dr.math. Jānis Valeinis

_____ _____

(paraksts) (datums)

Recenzents: lektors Jānis Smotrovs

_____ _____

(paraksts) (datums)

Darbs iesniegts Matematikas nodaļā _____

(datums)

_____

(darbu pieņēma)

Darbs aizstāvēts maģistra gala pārbaudījuma komisijas sēdē

_____ prot. Nr. _____, vērtējums_____

(datums)

Komisijas sekretārs/-e: _____ _____

(Vārds, Uzvārds) (paraksts)