Монгол хүний амьдрах хугацааны тархалтын функцийн В-сплайн үнэлэлт

О.Цэрэнбат[1], О.Гантөмөр[2], Б.Чулуунпүрэв[3]

[1] Монгол Улсын Их Сургууль, Математик Компьютерийн Сургууль
[2] Практикал даатгал ххк
[3] Монгол даатгал ххк

Удиртгал

Амьдрах хугацааны тархалтын функцыг зөв байгуулах асуудал нь тухайн улс үндэстний актуар тооцоонд чухал үүрэг гүйцэтгэдэг. Амьдрах хугацааны тархалтын функцийг

1. 1729 онд Де Муавра $0 \le x < w$

$$f(x) = \frac{1}{x}, F(x) = \frac{x}{w}, s(x) = 1 - \frac{x}{w}, \mu_x = \frac{1}{w - x}$$

2. 1825 онд Гомпертц $\mu_x = Be^{\alpha x}, \alpha > 0, B > 0$

$$s(x) = \exp[-B(e^{\alpha x} - 1)/\alpha], \ f(x) = B\exp[\alpha x - B(e^{\alpha x} - 1)/\alpha]$$

3. 1860 онд Мэйкхам $\mu_x = A + Be^{\alpha x}$

$$s(x) = \exp[-Ax - B(e^{\alpha x} - 1)/\alpha]$$

$$f(x) = [A + Be^{\alpha x}]\exp[-Ax - B(e^{\alpha x} - 1)/\alpha]$$

4. 1939 онд Вейбулл $\mu_x = kx^n$

$$s(x) = \exp[-kx^{n+1}/(n+1)]$$

$$f(x) = kx^{n}\exp[-kx^{n+1}/(n+1)]$$

гэх мэтээр таамаглан тархалтын параметрын утгыг үнэлж байсан боловч энэ таамаглал няцаагдсан юм.

Vladimir K Kaishev,Dimitrina S.Dimitrova нар амьдрах хугацааны тархалтын функцыг B-Сплайн аргаар үнэлэх аргачлалыг боловсруулсан.

Манай орны хувьд үндэстний хүн амын амьдрах хугацааны тархалтын функцийн талаар зарим нэг судалгаа сүүлийн үед хийгдэж байгаа боловч практикт хэрэглэх хэмжээнд хийгдсэн нь үгүй юм. Гэвч амьдрах хугацаа, насжилтын хүснэгтийн хэрэглээ нь нийгмийн даатгал, амьдралын даатгалд ихээхэн хэрэгтэй байгааг мэргэжлийн байгууллагууд ч хүлээн зөвшөөрч байна.

Энэхүү судалгааны ажлаараа Их Британийн насжилтын хүснэгтийг боловсруулсан аргыг суурь болгон ашиглаж Монгол хүний амьдрах хугацааны тархалтын функцын үнэлэлтийг байгуулах математик загварыг боловсруулан, MATLAB 2009 программ дээр үр дүнг шалгалаа. Ингэснээрээ олон улсад хүлээн зөвшөөрөгдсөн аргаар үндэсний хүн амын насжилтын хүснэгтийг байгуулах боломжтой болж байгаа юм. Судалгааны ажлыг боловсруулахдаа 2003-2008 оны нас баралтын тоон мэдээлэлийг өгөгдөл болгон ашиглав.

1. Амьдрах хугацааны тархалт

Хүний амьдрах хугацааг X гэе. $X \geq 0$ тасралтгүй, санамсаргүй хэмжигдэхүүн бөгөөд түүний тархалтын функцыг $F_X(x)$ гэж тэмдэглэе.

$$F_X(x) = P(X \leq x), \qquad x \geq 0$$

$s(x) = 1 - F_X(x) = P(X > x), \quad x \geq 0$ функцыг амьдралын функц гэнэ.

x настай хүн ойрын Δx хугацааны дотор нас барах нөхцөлт магадлал

$$P(x \leq X < x + \Delta x / X \geq x) = \frac{F(x+\Delta x) - F(x)}{1 - F(x)} = \frac{f(x)}{1 - F(x)}\Delta x + o(\Delta)$$

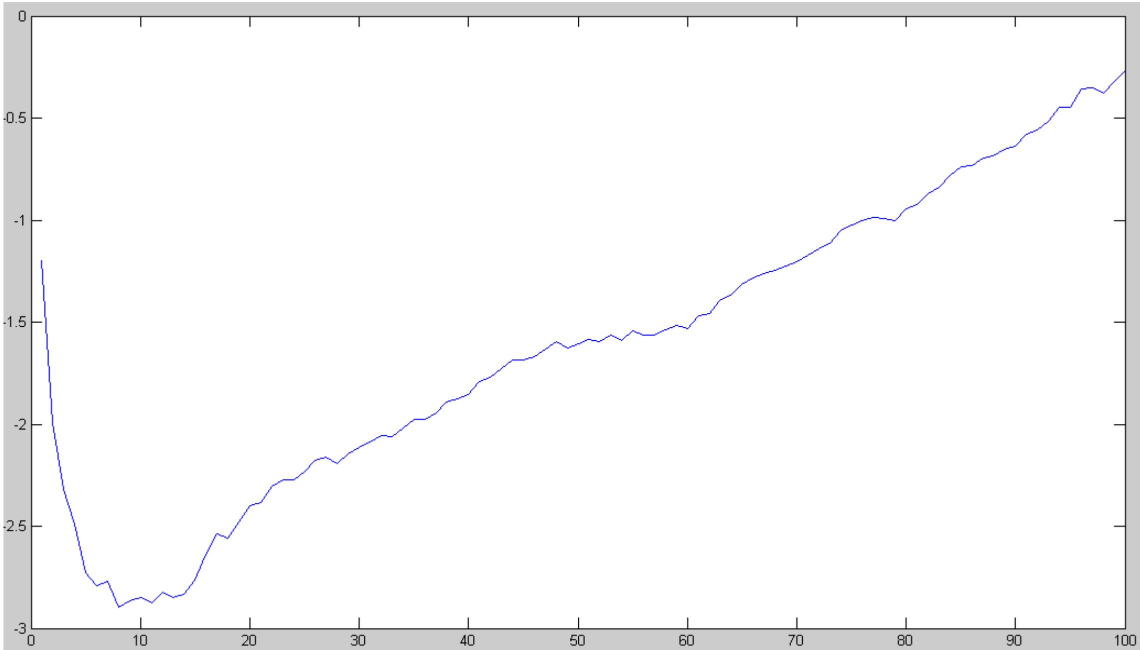
гэж тодорхойлогдно. Энд:

$f(x) = F'(x)$ - амьдрах хугацааны тархалтын нягтын функц ,

$\lambda(x) = \dfrac{f(x)}{1 - F(x)}$ - мөхлийн эрчим.

Амьдрах хугацааны тархалтын функц нь мөхлийн эрчмээр $F(x) = 1 - e^{-\int_0^x \lambda(t)dt}$ гэж илэрхийлэгдэнэ.

2. Амьдрах хугацааны тархалтын функцын эмперик үнэлэлт



$[x-1, x[$ насны интервалд нас барсан хүний тоог v_x гэж тэмдэглье. $x = 1, 2, \ldots, 100$

100 аас дээш насандаа нас барсан хүний тоог v_{101} гэж тэмдэглье.

$S = \sum_{x=1}^{101} v_x$ нийт нас барагсдын тоо.

$F(x)$ амьдрах хугацааны тархалтын функцын x цэг дээрх эмперик үнэлэлт нь

$$\hat{F}(x) = \frac{v_1 + v_2 + \ldots + v_x}{S} \quad x = 1, 2, \ldots, 101$$ байна.

Мөхлийн эрчимийн функцын x цэг дээрх эмперик үнэлэлт нь

$$\lambda(x) = \frac{\hat{F}(x) - \hat{F}(x-1)}{1 - \hat{F}(x)} \quad x = 1, 2, \ldots, 100$$ байна.

2003-2008 оны монгол хүний нас баралтын тоон мэдээллийг ашиглан $\lg \hat{\lambda}(x)$ эмперик үнэлэлтийн график байгуулбал.

3. B-сплайн муруйн тодорхойлолт

$$f(x) = \sum_{i=1}^{p} c_i N_{i,n}(x)$$

(1)

гэж тодорхойлогдсон (1)-г B-сплайн муруй гэж нэрлэх бөгөөд энд c_i-н хяналтын цэгүүд, $N_{i,n}(x)$ -н B-сплайны суурь функцууд, ***p***-н хяналтын цэгүүдийн тоо.

Зангилааны цэгүүдийн олонлог

$$t_{k,n} = \left\{ t_1 = \ldots = t_n = a < t_{n+1} < \ldots < t_{n+k+1} = \ldots t_{2n+k} = b \right\}$$

(2)

дээр тодорхойлогдсон Mansfield-De Boor-Cox-н

$$N_{i,1}(t) = \begin{cases} 1 & хэрэв \; t_i \le t < t_{i+1} \\ 0 & бусад \end{cases}$$

$$N_{i,n}(t) = \frac{t - t_i}{t_{i+n-1} - t_i} N_{i,n-1}(t) + \frac{t_{i+n} - t}{t_{i+n} - t_{i+1}} N_{i+1,n-1}(t)$$

(3)

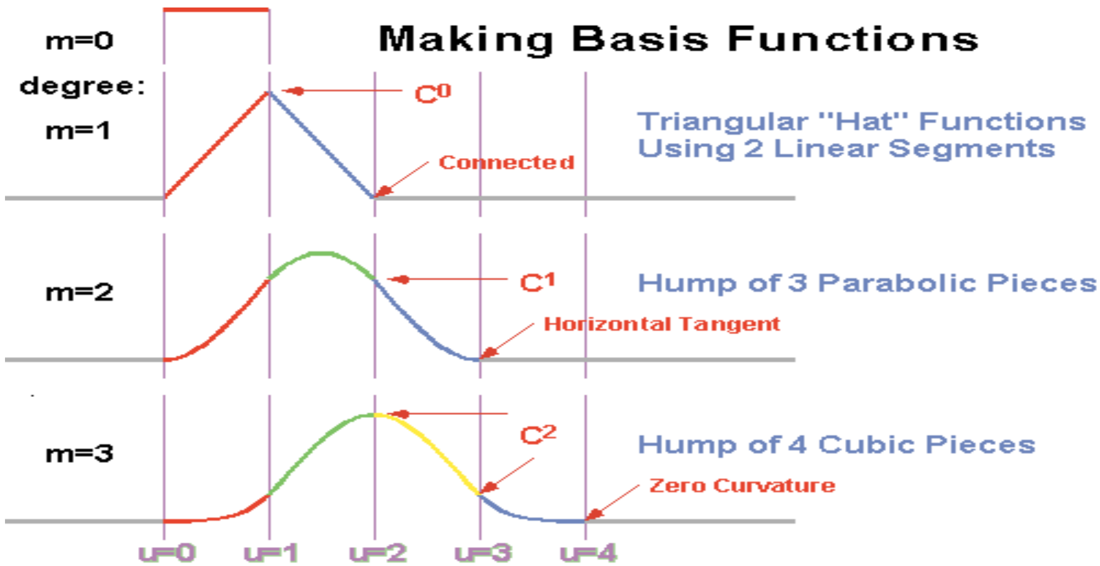
$$N_{i,n}(t) = 0 \qquad t \notin [t_i, t_{i+n}]$$

рекуррент томъёогоор тодорхойлогдсон $N_{i,n}(x)$ -г B-сплайны суурь функцууд гэнэ.

$N_{i,n}(x)$ -нь n-1 зэргийн хэсэг хэсэг интервал дээр тодорхойлогдсон олон гишүүнтүүдийн нийлбэр байна.

B-сплайн функцийн чанарууд:

- ✓ $N_{i,n}(x)$ -н $t \in [t_i, t_{i+n}]$ завсарт сөрөг биш утгатай функц байна.
- ✓ $\sum_{i=1}^{p} N_{i,n}(x) = 1$ $\qquad t \in [t_i, t_{i+n}]$ байна.
- ✓ n-1 эрэмбийн тасралтгүй дифференциалтай. Өөрөөр хэлбэл $N_{k,n}(t) \in C^{n-1}[a,b]$



4. Логарифмчилсан мөхлийн эрчмийн B-сплайн дөхөлтийн үнэлэлтийг байгуулах сплайн регрессийн арга

$\lg \hat{\lambda}(x)$ -нь мөхлийн эрчмийн ажиглалтын утга, $\lg \lambda(x)$ -н мөхлийн эрчмийн жинхэнэ муруй буюу манай хайж буй муруйг тэмдэглэе.

$$\lg \hat{\lambda}(x) = \lg \lambda(x) + e \qquad x \in [a, b] \tag{4}$$

байна. Энд e- нь тэг дундажтай санамсаргүй алдаа, $\lg \lambda(x)$ - н үл мэдэгдэх олон гишүүнт бөгөөд $\lg \lambda(x)$ -г n-1 зэргийн олон гишүүнт сплайн $f(t_{k,n}, x)$ -р дөхөе.

$f(t_{k,n}, x)$ - н

$$t_{k,n} = \{t_1 = \ldots = t_n = a < t_{n+1} < \ldots < t_{n+k+1} = b = \ldots t_{2n+k}\} \tag{5}$$

зангилааны цэгүүдийн олонлог дээр тодорхойлогдсон

$$f(t_{k,n}, x) = \theta' N_n(x) = \sum_{i=1}^{p} \theta_i N_{i,n}(x) \tag{6}$$

хэсэг хэсэг олон гишүүнт функц юм.

Энд $\theta = (\theta_1, \ldots, \theta_p)$ -н регрессийн коэффициентүүдийн вектор ба $N_n(x) = (N_{1,n}(x), \ldots, N_{p,n}(x))$, $p = n + k$, n-1 зэргийн B-сплайнууд. B-сплайнууд нь $t_{k,n}$ дээр тодорхойлогдсон Mansfield-De Boor-Cox-н томъёогоор тодорхойлогдох олон гишүүнт.

Сплайн регрессийн зорилго нь x насанд харгалзах мөхлийн эрчим $\{\lg \hat{\lambda}(x)\}_{x=1}^{N}$ -нь ажиглалтын утгуудад тулгуурлан сплайны зэрэг n-1, зангилааны цэгүүдийн тоо k, регрессийн коэффициент $\theta = (\theta_1, \ldots, \theta_p)$ - г үнэлэх юм. Энд N=100. Зангилааны векторыг үнэлсний дараа Хамгийн бага квадратын аргаар регрессийн коэффициент θ -г үнэлэх бөгөөд эндээс θ –н үнэлэлт

$\hat{\theta} = (F^T F)^{-1} F^T Y$ байна.

Энд F нь Nxp хэмжээст B-сплайн функцуудын матриц

$$F = \begin{pmatrix} N_{1,n}(1) & \ldots & N_{p,n}(1) \\ \vdots & \ddots & \vdots \\ N_{1,n}(100) & \cdots & N_{p,n}(100) \end{pmatrix}, \qquad Y = (y_1, \ldots, y_N)' ,$$

$y_i = \lg \hat{\lambda}(i) \quad i = 1, 2, \ldots, N$.

5. Параметрт сплайн регрессийн үнэлэлтийн геометр утга

(3)-д өгөгдсөн сплайн регресийн $Q(t)$ параметрт тохиолдолыг авч үзье.

$$Q(t) = \{x(t), y(t)\} = \left\{ \sum_{i=1}^{p} \xi_i N_{i,n}(t), \sum_{i=1}^{p} \theta_i N_{i,n}(t) \right\}$$

Энд: t нь параметр, $x(t)$ ба $y(t)$ нь $t_{k,n}$ зангилааны олонлог дээр тодорхойлогдсон сплайн функцууд. $x(t)$ -н

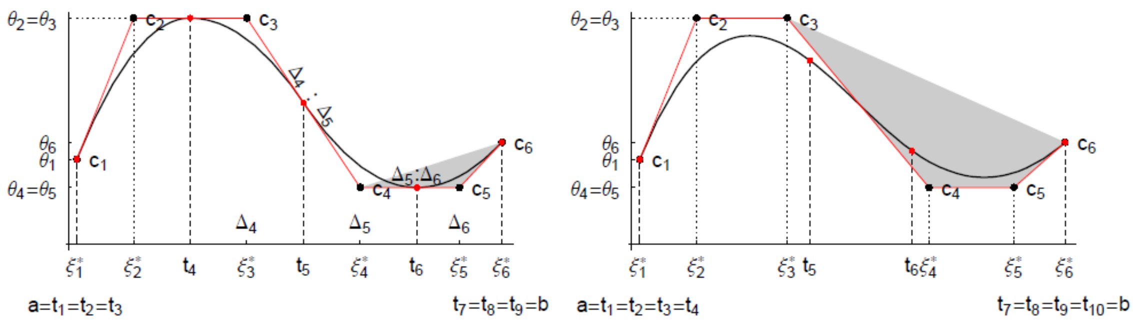
$$x(t) = \sum_{i=1}^{p} \xi^*_i N_{i,n}(t) = t \tag{8}$$

адитгалыг хангана. Энд ξ^* - нь n-1 ширхэг дараалсан зангилааны цэгүүдийн дунджаар тодорхойлогдоно. Өөрөөр хэлбэл:

$$\xi^* = (t_{i+1} + \ldots + t_{i+n-1}) / (n-1) \qquad i = 1, 2, \ldots p \qquad (9$$

$f(t_{k,n}, \theta : t) \quad t \in [a, b]$ - сплайн функцаар $Q(t)$ параметрт хэлбэрийг илэрхийлбэл

$$Q^*(t) = \{t, f(t_{k,n}, \theta : t)\} = \left\{ \sum_{i=1}^{p} \xi_i^* N_{i,n}(t), \sum_{i=1}^{p} \theta_i N_{i,n}(t) \right\} \qquad (10)$$



(9)-р өгөгдсөн ξ_i^* -г Grivelle –ийн абцисс гэнэ. $\xi^* = \{\xi_1^*, \xi_2^*, \ldots, \xi_p^*\} = \xi^*(t_{k,n})$

$Q^*(t)$ - дээр тодорхойлогдсон C_{Q^*} полигоныг хяналтын полигон гэх бөгөөд оройн цэгүүдийг нь хяналтын цэгүүд гээд C_i - ээр тэмдэглье. $C_i = (\xi_i^*, \theta_i)$ i=1,…,p

$Q^*(t)$ (n-1) - зэргийн сплайн регресийн муруйн цэг бүр ‘n’ ширхэг C_i хяаналтийн цэгийн
гүдгэр комбинац байна. Өөрөөр хэлбэл:

$$Q^*(t) = \sum_{i=j}^{n+j-1} c_i N_{i,n}(t) \qquad t \in [t_{n+j-1}, t_{n+k}] \quad \text{j=1,…,k}$$

$C_j, \ldots, C_{j+n-1}$ -ийн гүдгэр бүрхэвч гэж эдгээр цэгүүдэд хашигдсан жижиг полигоныг хэлнэ.

$C_i = (\xi_i^*, \theta_i)$ оройнуудтай C_{Q^*} хяналтын полигон нь шугаман сплайн функц ба дараах томъёогоор илэрхийлэгдэнэ.

$$C_{Q^*} = \left\{ \sum_{i=1}^{p} \xi_i^* N_{i,2}(t), \sum_{i=1}^{p} \theta_i^* N_{i,2}(t) \right\} = \left\{ t, \sum_{i=1}^{p} \theta_i^* N_{i,2}(t) \right\} = \sum_{i=1}^{p} \theta_i^* N_{i,2}(t) \qquad (11)$$

6. Мөхлийн эрчмийн B-сплайн муруйн зангилааны векторыг оновчлох нь

Энэ бүлэгт сплайн регрессийн зангилааны векторыг оновчлох аргыг товчхон авч үзье. Энэ арга нь сплайн регрессийн дотоод зангилааны цэгүүдийг хамгийн бага квадратын аргаар шугаман B-сплайн үнэлэх замаар оновчлох, Зангилааны вектор, сплайны зэргийг оновчлох үндсэн 2 алхамтай.

1-р алхамд (Stage A) Өгөгдсөн цэгүүдийг шулуун шугамаар холбож хамгийн бага квадратын аргаар үнэлэх замаар оновчтой зангилааны цэгүүдийг гарган авах тэдгээрийн тусламжтайгаар хяналтын полигоны анхны байрлалыг хадгалсан тохиромжтой хяналтын полигон байгуулах нь 1-р алхамын гол зорилго юм. 2-р алхамд 1-р алхамд олж авсан шугаман сплайны хяналтын цэгүүд дээр тулгуурлан максимум нормыг минимумчлах замаар зангилааны вектор, сплайны зэргийг оновчлох юм.

6.1 Хамгийн бага квадратын аргаар шугаман B-сплайн үнэлэх замаар дотоод зангилааны цэгүүдийг оновчлох 1-р шат

Нэг зэрэг зангилааны цэг нэмж болон шулуун шугамыг үнэлэхээс эхлэе.

$\sigma_{l,2} = \{\sigma_1 = \sigma_2 < \sigma_3 < \ldots < \sigma_{l+2} < \sigma_{l+3} = \sigma_{l+4}\}$ - зангилааны олонлогтой, p=l +2 ширхэг B-Сплайнтай , l – ширхэг дотоод зангилаатай $\hat{f}(\sigma_{l,2}, \hat{\alpha}, x)$ шугаман сплайныг хамгийн бага квадратаар хайя. Үлдэгдлийн квадратуудын нийлбэрийн харьцаа нь:

$$\frac{RSS(l+q)}{RSS(l)} = \frac{\sum_{j=1}^{N} (y_j - \hat{f}(\sigma_{l+q,2}, \hat{\alpha}, x_j))^2}{\sum_{j=1}^{N} (y_j - \hat{f}(\sigma_{l,2}, \hat{\alpha}, x_j))^2} \geq \alpha_{exit} \qquad (9)$$

Энд α_{exit} - босгын түвшин бөгөөд $\alpha_{exit} = 0.9$ байх нь хамгийн тохиромжтой.

(9) –нь q – ээс их зангилаа нэмэгдсэн тохиолдолд (q≥ 1) $\hat{f}(\sigma_{l,2}, \hat{\alpha}, x)$ - сайн дөхөлт болохгүй. Зангилааны цэг нэмсэн тохиолдолд $\hat{f}(\sigma_{l,2}, \hat{\alpha}, x)$ - г дахин үнэлнэ. Шугаман сплайнын үнэлэлт $\hat{f}(\sigma_{l,2}, \hat{\alpha}, x)$ нь $(\xi_i^*, \hat{\alpha}_i)$ - оройн цэгтэй хяналтын полигоноор илэрхийлэгднэ. Энд $\xi_i^* \equiv \sigma_{i+1}$ i=1,…,p байна.

$\hat{f}(\sigma_{l,2}, \hat{\alpha}, x)$ - нь Kaishev- ийн алгоритмаар байгуулагдана.

Энэ алгоритм нь хамгийн бага квадратын аргаар Шугаман B-сплайн регрессийг үнэлэх замаар зангилааны цэгийг оновчтой байрлалд, оновчтой тоогоор байрлуулах ба хяналтын цэгүүдийг байгуулна.

Алхам 1. n=2, k=0 үед буюу $t_{0,2} = \{a = t_1 = t_2 < t_3 = t_4 = b\}$ зангилааны олонлогтой үед эхэлье.

$f(t_{0,2}, \widehat{\alpha}, x) = \widehat{\alpha}_1 N_{1,2}(x) + \widehat{\alpha}_2 N_{2,2}(x)$ шугаман сплайны $\widehat{\alpha}$ коэффициентийн үнэлэлтийг хамгийн бага квадратын аргаар үнэлнэ.

$r_i = r(x_i) = y_i - \widehat{f}(t_{0,2}, \widehat{\alpha}, x_i)$ үлдэгдэлүүд болон үлдэгдэлүүдийн квадратуудын нийлбэр $RSS(k) = \sum_{i=1}^{N} r_i^2$ -г тооцоё.

Алхам 2: $r_i, i = 1, \ldots N$ үлдэгдэлүүдийн дарааллыг тэмдэгээр нь бүлэглэн кластеруудад хуваая. Өөрөөр хэлбэл $sign(r_1) = \ldots = sign(r_{d_1}) \neq sign(r_{d_1+1}) = sign(r_{d_{1+2}}) = \ldots = sign(r_{d_1+d_2}) \neq$

$\ldots \neq sign(r_{d_1+d_2+\ldots+d_{u-1}+1}) = sign(r_{d_1+d_2+\ldots+d_{u-1}+2}) = \ldots = sign(r_{d_1+d_2+\ldots+du})$

байх $1 \leq u \leq N$ байх u ба $d_j > 0$, $j = 1, \ldots, u$ бүхэл тоонууд d_j -г хайя.

Энд $\sum_{j=1}^{l} d_j = N$ байна.

Алхам 3: ижил тэмдэгтэй үлдэгдлүүдийн дарааллын кластер бүрийн хувьд үлдэгдэлийн дундаж m_j, $j = 1, \ldots, u$ утгуудыг тооцоё.

$$m_j = (\sum_{i=1}^{d_j} r_{d(j)+i}) / d_j = (\sum_{i=1}^{d_j} (\widehat{f}_{d(j)+i} - E\widehat{f}_{d(j)+i}) + (E\widehat{f}_{d(j)+i} - f_{d(j)+i}) + e_{d(j)+i}) / d_j =$$

$$= \sum_{i=1}^{d_j} (\widehat{f}_{d(j)+i} - E\widehat{f}_{d(j)+i}) / d_j + \sum_{i=1}^{d_j} (E\widehat{f}_{d(j)+i} - f_{d(j)+i}) / d_j + \sum_{i=1}^{d_j} e_{d(j)+i} / d_j \quad j = 1, \ldots, u,$$

сүүлийн задаргааны 3-н нийлбэр нь харгалзан дисперс, хазайлт,алдааг илэрхийлнэ. Энд

$d(j) = d_1 + d_1 + \ldots + d_{j-1}$ мөн ижил тэмдэгтэй үлдэгдлүүдэд харгалзах кластеруудын x-аргументын мужийг байгуулан хоорондох зайг нь олно. Өөрөөр хэлбэл:

$[x_{d(j)+1}, x_{d(j+1)}]$ мужуудыг байгуулж $\xi_j = x_{d(j+1)} - x_{d(j)+1}$, $j = 1, \ldots, u$

Алхам 4. $m_{\max} = \max_{1 \leq j \leq u}(m_j)$ ба $\xi_{\max} = \max_{1 \leq j \leq u}(\xi_j)$ -уудын тусламжтайгаар m_j, ξ_j утгуудыг нормчлоё. Өөрөөр хэлбэл $m_j{}' = m_j / m_{\max}$ ба $\xi_j{}' = \xi_j / \xi_{\max}$ харгалзан $0 < m_j{}' \leq 1$, $0 < \xi_j{}' \leq 1$ утгуудыг гаргах.

Алхам 5: w_j, $= j$ 1 кластеруудын жинг тооцно. $w_j = \beta m'_j + (1-\beta)\xi'_j$, $j = 1,\ldots,u$ Энд β параметр нь $0 \le \beta \le 1$ байна.

Алхам 6: w_j, $j = 1,\ldots u$ жингүүдийн буурах эрэмбийн дагуу кластеруудын дараалллаа эрэмбэлээд үүссэн эрэмбийн индексээр шинэ дараалал үүсгэе. Өөрөөр хэлбэл: $w_{j_1} \ge w_{j_2} \ge \ldots \ge w_{j_l}$ гэдгээс $\{j_1, j_2, \ldots, j_l\}$ дараалал үүснэ. Дараагийн алхамд j_1-р кластерт харгалзах 3-р алхамд үүсгэсэн x-н утгуудын интервалд шинэ зангилаа нэмэх замаар $\widehat{f}(t_{k,2}, \widehat{\alpha}, x)$ дөхөлтийг сайжруулая.

Алхам 7: к=0 байхад Хамгийн өндөр жинтэй j_1-р кластерт зангилааны нэг цэг нэмнэ. $k \ge 1$ байхад j_1-р кластераас эхлэн $j_2,\ldots j_s$ зангилааны цэг байгаа эсэхийг шалган зангилааны цэг агуулаагүй кластерт дараах томъёогоор цэг нэмнэ.

$$t^* = \Big(\sum_{i=d(j_s)+1}^{d(j_s)+d_{j_s}} r_i x_i\Big) / \Big(\sum_{i=d(j_s)+1}^{d(j_s)+d_{j_s}} r_i\Big)$$

Эндээс $t^* \in [x_{d(j_s)+1}, x_{d(j_s)+d_{j_s}}]$ байх цэг нэмэгдэх ба $t^*_{k+1,2} := t_{k,2} \cup \{t^*\}$ болно.

Алхам 8: Шинээр байгуулагдсан $t^*_{k+1,2}$-г ашиглан хамгийн бага квадратын аргаар дахин шугаман сплайнаа үнэлнэ.

$$\widehat{f}(t^*_{k+1,2}, \widehat{\alpha}, x) = \sum_{i=1}^{p} \widehat{\alpha}_i N_{i,2}(x)$$

Өөрөөр хэлбэл зангилаа нэмэгдэх бүрт **p**-н нэгээр нэмэгдэнэ.

Алхам 9: $\widehat{f}(t^*_{k+1,2}, \widehat{\alpha}, x)$ үед $RSS(k+1)$ ба r_i, $i = 1,\ldots N$ -г тооцно. Хамгийн бага квадратын аргын үнэлэлтийн ортогналь чанараар

$$\sum_{i=1}^{N}(y_i - \widehat{f}(t_{k,2}, \widehat{\alpha}, x_i))^2 = \sum_{i=1}^{N}(y_i - \widehat{f}(t^*_{k+1,2}, \widehat{\alpha}, x))^2 + \sum_{i=1}^{N}(\widehat{f}(t^*_{k+1,2}, \widehat{\alpha}, x) - \widehat{f}(t_{k,2}, \widehat{\alpha}, x_i))^2$$

Эндээс $RSS(k+1) < RSS(k)$ гэдэг нь илэрхий юм.

Алхам 10: $q \ge 1$ байх бэхлэгдсэн тоо байг. 1-р шат эхлэхэд сонгогдсон байг. $t^*_{k+1,2}$ зангилааны цэгүүдийн олонлгийн дотоод зангилааны цэгүүдийн тоо нь q-ээс бага бол 2-р шатруу шилжин үгүй бол (q-ээс олон) $\alpha = RSS(k+1) / RSS(k+1-q)$ харьцааг тооцоё. 9-р шатнаас $0 < \alpha < 1$

гэдэг нь харагдана. Хэрэвээ $\alpha \geq \alpha_{exit}$ бол $\widehat{f}(t_{l,2}, \widehat{\alpha}, x_i),\ l = k+1-q$ үнэлэлтийг аваад А шатнаас гарна. Хэрвээ $\alpha \leq \alpha_{exit}$ бол $\widehat{f}(t^*_{k+1,2}, \widehat{\alpha}, x)$ үнэлэлтийг аваад дахин 2-р алхам руу очно.

Энэ 10 алхам нь давтан хийгдсээр хамгийн оновчтой зангилааны цэгийн олонлог буюу хяналтын полигоныг бидэнд өгөх юм.

6.2 Зангилааны вектор, сплайны зэргийг оновчлох 2-р шат:

$n = 3, \ldots, n_{\max}$ - ийн утгуудын хувьд дараах минимумчлах бодлогын шийдийн тусламжтайгаар $\tilde{t}_{l-(n-2),n}$ зангилааны цэгүүдийг оновчлоё.

$$\min_{\substack{t_{l-(n-2),n} \\ \xi_{i+1} < t_{i+n} < \xi_{i+n-1} \\ i=1,\ldots,k}} \left\| \widehat{f}(\sigma_{l,2}, \widehat{\alpha}, x) - C_{f(t_{l-(n-2),n}, \widehat{\alpha}, x)} \right\| \qquad (13)$$

Энд $\|g\| := \max_{a \leq x \leq b} |g(x)|$.

(13) минимумчлах бодлогын шийдийн тусламжтайгаар 1-р шатанд үнэлсэн хяналтын цэг $\widehat{\alpha}_i$ - ээр у координат, $\xi(\tilde{t}_{l-(n-2),n})$ Grivelle-абциссаар х координатаа хийсэн полигон байгуулж болно. Энд байгуулсан $\widehat{f}(\sigma_{l,2}, \widehat{\alpha}, x)$ ба $C_{f(t_{l-(n-2),n}, \widehat{\alpha}, x)}$ 2 полигон нь ижил l +2 ширхэг оройн цэгтэй боловч дотоод зангилааны цэгүүдийн тоо нь харгалзан l , l -(n-2) ширхэг юм.

Хяналтын цэгийн оновчлол, B-сплайны байгуулалт

Нэгэнт дотоод зангилааны тоо $\hat{l} - (\hat{n} - 2)$, сплайны зэрэг $\hat{n} - 1$, зангилааны вектор $\tilde{t}_{\hat{l}-(\hat{n}-2),\hat{n}}$, -г үнэлсний дараа B-сплайны хяналтын цэгүүдийг $\{x_i, y_i\}_{i=1}^{N}$ өгөгдлүүдийг ашиглан хамгийн бага квадратын аргаар θ үл мэдэгдэх параметрийг үнэлье.

$$\min_{\theta} \left[\sum_{i=1}^{N} (y_i - f(t_{l-(n-2),n}, \theta, x_i))^2\right]$$

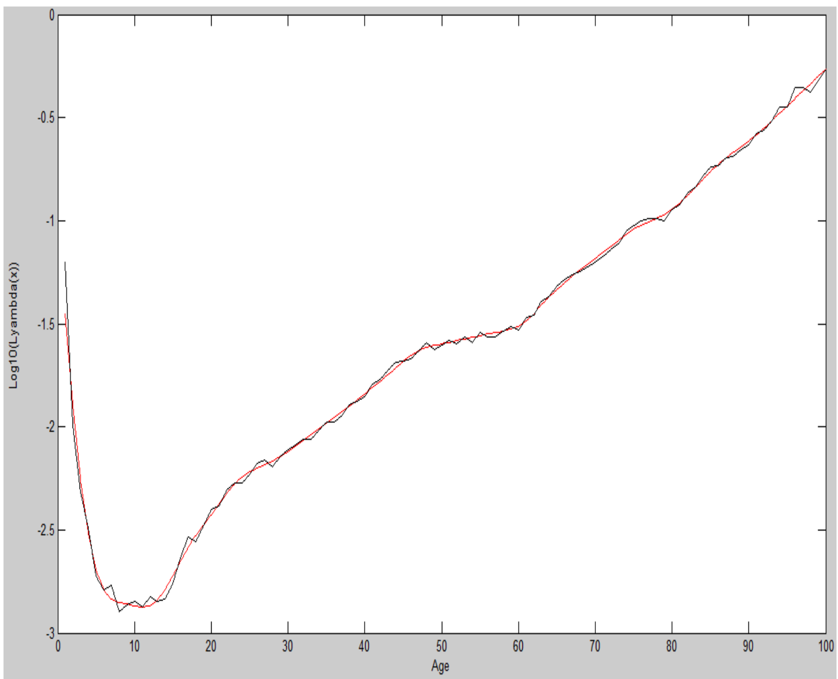
Минимумчлах бодлогыг шийд нь

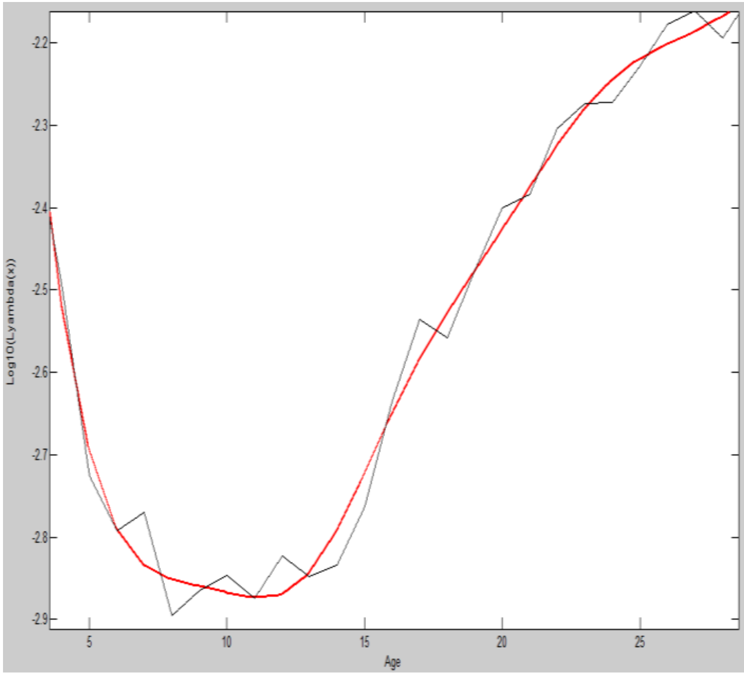
$$\hat{\theta} = (F^T F)^{-1} F^T Y$$

болох ба энд F-н Nxp хэмжээст B-сплайн функцуудын матриц i=1,…,N

$N_n(x) = (N_{1,n}(x), \ldots, N_{p,n}(x))'$, x_i i=1,…,N, $Y = (y_1, \ldots, y_N)'$

$f(\tilde{t}_{l-(\tilde{n}-2),\tilde{n}}, \hat{\theta}, x)$ B-сплайнаа байгуулж lg $\lambda(x)$ функцийг дөхөх зорилгодоо хүрлээ. B-сплайнаар дөхсөн мөхлийн эрчмийн графикийг, өгөгдлийн графиктай харьцуулан доорх зурагт харуулав. Энэ нь нийт хүн амын хувьд үнэлсэн мөхлийн эрчим юм.





Судалгааны үр дүнд дотоод зангилааны тоо $\tilde{k} = 20$, сплайны зэрэг $\tilde{n} - 1 = 2$ буюу квадрат сплайн гэж гарсан бөгөөд зангилааны утгуудын олонлог нь

$\tilde{t}_{20,3}$ ={ 1, 1, 1, 1.779, 2.6373, 3.56561, 4.6969, 5.74059,6.4835, 12.7378, 17, 18, 23.69244, 28, 38.631, 46.988, 61.193, 62, 65.613, 75.1796, 79.9025, 86, 90.308, 100 , 100 , 100 }

гэж олж гаргасан. Эндээс бэхлэгдсэн x насны хувьд мөхлийн эрчмийн логарифм нь

$$\lg \hat{\lambda}(x) = c_{i,0} + c_{i,1}x + c_{i,2}x^2 \qquad t_{3+i} \le x < t_{3+i+1} \qquad i = 1,\dots 20$$ квадрат

тэгшитгэлээр олдоно. Энд $c_{i,0}, c_{i,1}, c_{i,2}$ -н хамгийн бага квадратын аргаар үнэлсэн коэффициентүүд. Эндээс мөхлийн эрчмийн үнэлгээг бүрэн хийж болох бөгөөд энэхүү хэсэг хэсэг квадрат сплайны $c_{i,j}$ коэффициентийг θ_i хяналтын цэгүүд, B-сплайн функцийн задаргаанаас үнэлбэл:

$$c_{i,0} = \begin{cases} \theta_i \cdot \dfrac{1}{(t_{i+2} - t_i)(t_{i+1} - t_i)} \cdot t_i^2, & t_i \le t < t_{i+1} \\ -(\theta_i \cdot \dfrac{t_{i+2} \cdot t_i}{(t_{i+2} - t_i)(t_{i+2} - t_{i+1})} + \theta_i \cdot \dfrac{t_{i+1} \cdot t_{i+3}}{(t_{i+3} - t_{i+1})(t_{i+2} - t_{i+1})}), & t_{i+1} \le t < t_{i+2} \\ \theta_i \cdot \dfrac{1}{(t_{i+3} - t_{i+1})(t_{i+3} - t_{i+2})} \cdot t_{i+3}^2, & t_{i+2} \le t < t_{i+3} \end{cases}$$

$$c_{i,1} = \begin{cases} -\theta_i \cdot \dfrac{2}{(t_{i+2} - t_i)(t_{i+1} - t_i)} \cdot t_i, & t_i \le t < t_{i+1} \\ \theta_i \cdot \dfrac{t_{i+2} + t_i}{(t_{i+2} - t_i)(t_{i+2} - t_{i+1})} + \theta_i \cdot \dfrac{t_{i+1} + t_{i+3}}{(t_{i+3} - t_{i+1})(t_{i+2} - t_{i+1})}, & t_{i+1} \le t < t_{i+2} \\ \theta_i \cdot \dfrac{-2}{(t_{i+3} - t_{i+1})(t_{i+3} - t_{i+2})} \cdot t_{i+3}, & t_{i+2} \le t < t_{i+3} \end{cases}$$

$$c_{i,2} = \begin{cases} \theta_i \cdot \dfrac{1}{(t_{i+2} - t_i)(t_{i+1} - t_i)} \quad , & t_i \le t < t_{i+1} \\ -(\theta_i \cdot \dfrac{1}{(t_{i+2} - t_i)(t_{i+2} - t_{i+1})} + \theta_i \cdot \dfrac{1}{(t_{i+3} - t_{i+1})(t_{i+2} - t_{i+1})}), & t_{i+1} \le t < t_{i+2} \\ \theta_i \cdot \dfrac{1}{(t_{i+3} - t_{i+1})(t_{i+3} - t_{i+2})}, & t_{i+2} \le t < t_{i+3} \end{cases}$$

болно. Мөхлийн эрчмийн үнэлгээг бүрэн хийж гүйцэтгэснээр амьдрах хугацааны тархалт, амьдралын хүснэгтийг байгуулах бүрэн боломжтой боллоо.

Энэхүү В-сплайн үнэлэлтийг Колмогоров-Смирновын шинжүүрээр $\alpha = 0.05$ итгэх түвшинд шалгаж үзсэн бөгөөд тэг таамаглалыг дээрх итгэх түвшинд няцаахгүй.

Насны интервал		Коэффициент - Эр			Насны интервал		Коэффициент - Эм		
		c1	c2	c3			c1	c2	c3
1.00	1.95	-0.96	1.91	-0.96	1.00	1.79	-1.42	2.83	-1.42
1.95	2.94	-1.12	4.37	-4.27	1.79	2.63	-1.40	4.99	-4.44
2.94	3.88	-1.62	9.57	-14.10	2.63	3.56	-1.27	6.70	-8.84
3.88	4.50	-0.45	3.50	-6.81	3.56	4.69	-1.08	7.72	-13.77
4.50	13.40	-0.02	0.20	-0.46	4.69	5.74	-1.48	13.92	-32.70
13.40	19.00	-0.02	0.20	-0.46	5.74	6.48	-0.55	6.28	-18.02
19.00	23.80	-0.02	0.20	-0.46	6.48	12.73	-0.04	0.57	-1.84
23.80	29.00	-0.02	0.20	-0.46	12.73	17.00	-0.12	2.93	-18.69
29.00	33.50	-0.03	0.77	-5.15	17.00	18.00	-0.38	12.83	-109.07
33.50	35.60	-0.01	0.28	-2.66	18.00	23.69	-0.04	1.42	-12.77
35.60	39.00	-0.01	0.28	-2.66	23.69	28.00	-0.03	1.60	-18.94
39.00	48.00	0.00	0.24	-3.46	28.00	38.63	-0.01	0.52	-7.32
48.00	49.90	0.00	0.24	-3.46	38.63	46.98	-0.01	0.67	-12.86
49.90	52.50	0.00	0.48	-12.06	46.98	61.19	-0.01	0.66	-15.55
52.50	58.10	0.00	0.48	-12.06	61.19	62.00	-0.41	49.77	-152.83
58.10	64.90	0.00	0.48	-12.06	62.00	65.61	-0.03	3.41	-105.66
64.90	70.00	-0.01	1.19	-41.79	65.61	75.17	-0.01	0.99	-32.59
70.00	80.00	-0.01	1.19	-41.79	75.17	79.90	-0.02	2.81	-105.70
80.00	81.00	-0.11	17.67	-706.95	79.90	86.00	-0.01	1.82	-72.76
81.00	88.40	-0.01	0.82	-33.19	86.00	90.30	-0.01	1.74	-74.68
88.40	100.00	0.00	0.59	-26.03	90.30	100.00	0.00	0.51	-23.12

Ном зүй

[1] Geometrically Designed, Variable Knot RegressionSplines : Asymptotics and Inference.October 2006. Cass Business School, 106 Bunhill Row, London EC1Y 8TZ, Vladimir K Kaishev,Dimitrina S.Dimitrova,Steven Haberman,Richard J. Verrall

[2] Geometrically Designed, Variable Knot Regression Splines : Variation Diminishing Optimality of Knots. October 2006. Cass Business School, 106 Bunhill Row, London EC1Y 8TZ, Vladimir K Kaishev, Dimitrina S.Dimitrova, Steven Haberman, Richard J. Verrall

[3] Automatic, Computer Aided Geometric Design of Free-Knot, Regression Splines. August 2004. Cass Business School, 106 Bunhill Row, London EC1Y 8TZ, Vladimir K Kaishev, Dimitrina S.Dimitrova, Steven Haberman, Richard J. Verrall

[4] B-Spline Interpolation and Approximation. December 2006. State Key Lab of CAD&CG, Zhejiang University, Hongxin hang, JieqingFeng

[5] Spline Methods Draft May 2008. Department of Informatics Centre of Mathematics for Applications,University of Oslo,May 19, 2008 Tom Lyche, Knut Morken

[6] ACTUARIAL MATHEMATICS 1997, NEWTON L. BOWERS, JR.HANS U.GERBER,JAMES C. HICKMAN,DONALD A. JONES,CECIL J. NESBITT