BUTSTREP USULIDA QURILGAN STATISTIK BAHOLAR UCHUN LIMIT TEOREMALAR

Muradov R.
Namangan muhandislik-texnologiya instituti
Ziyoitdinova M.
Andijon davlat universiteti 3-bosqich tayanch doktoranti

Tayanch so'zlar: matematik statistika, butstrep usuli, statistik baho, ishonch oraliqlari, empirik taqsimot, limit teoremalar.

Ключевые слова: математическая статистика, бутстреп-метод, статистическая оценка, доверительные интервалы, эмпирическое распределение, предельные теоремы.

Key words: mathematical statistics, bootstrap method, statistical estimation, confidence intervals, empirical distribution, limit theorems.

Kirish va masalaning qo'yilishi. Ma'lumki, matematik statistikada ko'pgina baholash usullari, masalan, haqiqatga maksimal o'xshashlik usuli, momentlar usuli, xi-kvadrat minumimi usuli va boshqalar yordamida tuzilgan statistik baholarning aniq yoki limit taqsimotlari bosh to'plamning noma'lum taqsimotiga bog'liq bo'lib qolib, ularni amalda qo'llanilishida bir qancha qiyinchiliklarga olib keladi. Ammo butstrep baholar bunday kamchiliklardan holi bo'lib, ularni masalan noma'lum xarakteristikalar uchun ishonch oraliqlarini qurishda keng qo'llash mumkindir. Bundan tashqari, ular yordamida qurilgan ishonch oraliqlari berilgan ishonch ehtimolini o'zgartirilmagan holda odatdagi standart ishonch oraliqlaridan bir muncha torroq bo'lar ekan.

Noma'lum F taqsimotga ega bo'lgan bo'sh to'plamdan $X^{(n)}=(X_1, . . . ,X_n)$- statistik tanlanma olingan bo'lsin. Bu yerda $\{X_i\ ,\ i=\overline{1.n}\}$ bog'liq bo'lmagan va bir xil taqsimlangan tasodifiy miqdorlardir. Matematik statistikaning muhim masalalardan biri $X^{(n)}$ bo'yicha F ning biror funksionali $\theta=T(F)$ ni baholash masalasidan iboratdir. Ushbu maqolada biz θ- bir o'lchovli skalyar bo'lgan holni ko'ramiz. Demak, $\theta\in\Theta\subseteq R^1$. Agar F taqsimotlar oilasi noma'lum parametrlar aniqligida berilgan bo'lib, θ aynan mana shu parametrning o'zi yoki uning biror funksiyasi bo'lsa, u holda ko'rilayotgan masala parametrik, aks holda noparametrik deb ataladi.

Agar $\theta_n =У_n(X^{(n)})$ statistika θ ning biror bahosi bo'lsa, odatda θ_n ning sifati (xossalari) siljish kattaligi $B_n(F) =M_F(\theta_n-\theta)$, kvadratik risk $D_n(F) =M_F(\theta_n-\theta)^2$ yoki biror bir boshqa xarakteristikalar orqali aniqlanishi mumkin. Ammo ko'p hollarda bahoning xossalarini ifodalovchi bu kabi xatoliklar θ_n statistikaning quyidagi noma'lum

$$G_n(x;F) =P_F\{x^{(n)}: \theta_n(x^{(n)})<x\}$$

taqsimotga bog'liq bo'lib qolib, ulardan foydalanishda qiyinchiliklar yuzaga keladi. Bu yerda P_F, M_F ва D_F lar $X^{(n)}$ tanlanmani elementlari F ga teng degan shartda hisoblangan ehtimollik, matematik kutilma va dispersiyalarni anglatadi. Hozirgina ta'kidlaganimizni tasdiqlovchi bir misol keltiraylik.

1.1-misol. Kuzatilayotgan tasodifiy miqdor X ning taqsimoti F va matematik kutilmasi $\theta = M_F X$ bo‘lsin. Agar $X^{(n)} = (X_1, ..., X_n)$ tanlanma X ni kuzatish natijasida olingan bo‘lsa, u holda θ uchun tabiiy, ya’ni momentlar usuli bahosi

$$\theta_n = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

o‘rta arifmetik qiymatdir. U holda

$$M_F \overline{X} = \frac{1}{n} M_F \sum_{i=1}^{n} X_i = \frac{1}{n} \cdot n M_F X = \theta,$$

$$D_F \overline{X} = \frac{1}{n^2} D_F \left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \cdot n D_F X =$$

$$= \frac{1}{n} M_F (X-\theta)^2 = \frac{1}{n}\int_{-\infty}^{\infty} (x-\theta)^2 dF(x) = \frac{\sigma_F^2}{n} \qquad (1.1)$$

Demak, (1.1) hisoblar asosida

$$B_n(F) = M_F(\theta_n - \theta) = M_F \theta_n - \theta = \theta - \theta = 0$$

$$D_F(F) = M_F(\theta_n - \theta)^2 = M_F(\theta_n - M_F \theta_n)^2 = D_F \overline{X} = \frac{\sigma_F^2}{n}$$

Agar $\sigma_F^2 \in (0, \infty)$ bo‘lsa, u holda markaziy limit teoremasiga asosan yetarlicha katta n larda:

$$P_F\left(\frac{\sqrt{n}(\overline{X} - \theta)}{\sigma_F} < x\right) = P\left(\overline{X} < \theta + \frac{\sigma_F \cdot x}{\sqrt{n}}\right) = G_n\left(\theta + \frac{\sigma_F \cdot x}{\sqrt{n}}; F\right) \approx$$

$$\approx \Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$ - standart normal qonun.

Bu munosabatdan

$$G_n\left(\theta + \frac{x}{\sqrt{n}}; F\right) \approx \Phi(\sigma_F^{-1} \cdot x),$$

ya’ni $\theta_n = \overline{X}$ bahoning taqsimoti nomalum F ga bog‘liq bo‘lib qolmoqda.

Demak, yuqorida ko‘rib o‘tkanimiz kabi holatlarda noma’lum θ uchun θ_n yordamida ishonch oraliqlari qurish yoki biror gipotezalarni tekshirish uchun kriteriylarni qurishda muammolar yuzaga kelib qoladi. Amaliy statistikada uchraydigan bunday muammolarni hal qilishda θ uchun butsrep baholar qulaydir. Endi biz butsrep baholarini qurish usullari bilan tanishib o‘tamiz. Baholar qurishning butstrep usulini eng dastlab Bred Efron 1979-yilda [1,2,4] tavsiya etgan. Butstrepning mohiyati quyidagidan iboratdir.

Asosiy masala. (Ω, A, Q) ehtimollik fazosida aniqlangan kuzatilayotgan X tasodifiy miqdor yaratgan ehtimollik fazosini (X, B, P) orqali belgilaymiz. Bu yerda X - to‘plam X ning barcha mumkin bo‘lgan qiymatlari to‘plami, B esa X ning to‘plam ostilaridan tuzilgan Borel to‘plamlari $\sigma -$ algebrasi va X ning taqsimot qonuni

$$P(B) = Q(\omega: X(\omega) \in B), \quad B \in B,$$

Biror noma’lum parametr θ aniqligida berilgan (parametrik hol) yoki umuman (noparametrik hol) P oilaga tegishlidir, ya’ni

$$\{P_\theta, \theta \in \Theta\} = P \text{ ёки } \{P\} = P$$

$\hat{P}_n(\cdot)$ orqali $P(\cdot)$ ning biror bahosini belgilaymiz.

Masalan, parametrik holda

$$\hat{P}_n(B) = P_{\theta_n}(B), \quad B \in B,$$

bo‘lib, bu yerda θ_n statistika θ ning biror bahosi va noparametrik holda esa

$$\hat{P}_n(B)=\frac{1}{n}\sum_{k=1}^{n} I(X_k \in B),\quad B\in \mathcal{B}, \tag{1.2}$$

- emperik baho bo'lsin. Bu yerda I(A) orqali A xodisaning indikatorini belgiladik. Agar B=(-∞,x) deb olsak, P(B)=P((-∞, t))=F(t) – taqsimot funksiya bo'ladi va unga mos bahoni $\hat{P}_n(B)=\hat{P}_n((-\infty,t))=\hat{F}(t)$ deb belgilaymiz. Hajmi m bo'lgan $X_1^{(n)},\ldots,X_m^{(n)}$ tasodifiy miqdorlarni $X^{(n)}=(X_1,\ldots,X_n)$ tanlanma berilganligi shartidagi taqsimot qonuni $\hat{P}_n(\cdot)$ bo'lsin. Bu esa o'z navbatida quyidagini anglatadi, bir ehtimollik bilan

$$Q(\omega:\ X_1^{(n)}(\omega)\in B_1,\ldots, X_m^{(n)}(\omega)\in B_m / X^{(n)})=\prod_{k=1}^{m}\hat{P}_n(B_k),\quad B_k\in \mathcal{B},\ k=\overline{1.m} \tag{1.3}$$

(1.3) tenglik bilan aniqlanadigan $(X_1^{(n)},\ldots,X_m^{(n)})=Y_m^{(k)}$ statistik tanlanma <u>butstrep tanlanma</u> deb ataladi. Odatda m=m(n) va n tanlanma hajmlaridan

$$0<\liminf_{n\to\infty}\left(\frac{m}{n}\right)\le \limsup_{n\to\infty}\left(\frac{m}{n}\right)<\infty \tag{1.4}$$

shart talab qilinadi [1,4] (masalan m=n deb tanlash mumkin). Butstrep tanlanmasining taqsimot funksiyasi $\hat{F}_n(t)$ bo'lganligi sababli $Y_m^{(k)}$ ni statistik modellash usulida tuzish mumkin. $\theta^*_{nm}=\Psi_{nm}(X_1^{(n)},\ldots,X_m^{(n)})$ orqali θ uchun butstrep bahoni belgilab olamiz. θ^*_{nm} ning taqsimot funksiyasini $G^*(y)=P(y_m^{(n)}:\theta^*_{nm}<y/X^{(n)})$ orqali belgilab olib, uni baholashni quyidagi Efron taklif etgan uchta qadamlarda amalga oshiramiz:

<u>1-qadam</u>: $X^{(n)}$ tanlanma orqali $\hat{F}_n(t)$ bahoni qurib olamiz;

<u>2-qadam</u>: Berilgan m=1,...M lar uchun $\hat{F}_n$ taqsimotga ega bo'lgan $(X_1^{(n)},\ldots,X_m^{(n)})=Y_m^{(k)}$ butstrep tanlanmalarni va ular orqali esa θ uchun θ^*_{nm} butstrep baholarni qurib olamiz;

<u>3-qadam</u>: G^* ni baholash uchun $Y_m^{(n)}$ orqali emperik bahoni quramiz:

$$G_M^*(y)=\frac{1}{M}\sum_{m=1}^{M} I(\theta^*_{nm}<y)$$

Biz M^* orqali G^* taqsimotga nisbatan hisoblangan matematik kutilmani belgilab olib,

$$B_n^*=M^*\theta_{nm}^*-\theta_n\ ,$$

$$D_n^*=D^*\theta_{nm}^*=M^*(\theta_{nm}^*-M^*\theta_{nm}^*)^2,$$

lar orqali θ_{nm}^* butstrep bahoning siljish kattaligi va dispersiyalarini kiritamiz. Bu xarakteristikalarni G_m^* empirik taqsimot yordamida baholaymiz:

$$B_{nm}^*=\frac{1}{M}\sum_{m=1}^{M}\theta^*_{nm}-\theta=\overline{\theta_{nm}}-\theta_n,$$

$$D_{nm}^*=\frac{1}{M}\sum_{m=1}^{M}(\theta^*_{nm}-\overline{\theta_{nm}})^2.$$

Butstrep baholarning odatdagi baholarga nisbatan afzalliklari ko'pgina mualliflar tomonidan ta'kidlab o'tilgandir[1,2,4]. Bunday baholarning afzalligini ko'rsatish uchun ularni noma'lum parametrlar uchun ishonch oraliqlari qurish masalalarida namoyish etamiz. Matematik statistikadan malumki [1-4], agar noma'lum θ parametr uchun oddiy $X^{(n)}=(X_1,\ldots,X_n)$ tanlanma yordamida qurilgan θ_n baho berilganida θ_n- θ farq taqsimoti aniq va u F ga bog'liq bo'lmasa u holda θ uchun ishonch oralig'ini qurish qiyin emas. Haqiqatan, agar θ_n- θ ning aniq taqsimoti normal taqsimot $N(0,\sigma_n{}^2)$ va bunda $\sigma_n{}^2$ ma'lum (ya'ni F ga bog'liq emas) bo'lsa

$$P(\frac{(\theta_n-\theta)}{\sigma_n}<z)=\Phi(z), \qquad z\in R,$$

u holda standart normal taqsimot F(z) ning α va 1-α satxlaridagi ishonch ehtimollariga mos z_α va $z_{1-\alpha}$ kvantillari uchun

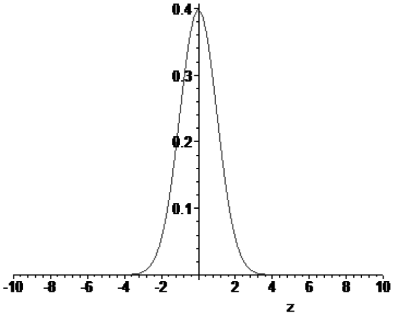
($\Phi(z_\alpha)=\alpha$, $\Phi(z_{1-\alpha})=1-\alpha$, $z_{1-\alpha}=z_\alpha$):

$$P(\theta<\theta_n+z_\alpha\cdot\sigma_n)=P(\frac{\theta_n-\theta}{\sigma_n}>-z_\alpha)=1-\Phi(-z_\alpha)=\alpha,$$

va

$$P(\theta<\theta_n+z_{1-\alpha}\cdot\sigma_n)=P(\frac{\theta_n-\theta}{\sigma_n}>-z_{1-\alpha})=1-\Phi(-z_{1-\alpha})=1-\Phi(z_\alpha)=1-\alpha,$$

munosabatlar o'rinli bo'lib, θ uchun izlanayotgan markaziy ishonch oralig'i



1-rasm.

$(\theta_n+z_\alpha\cdot\sigma_n;\theta_n+z_{1-\alpha}\cdot\sigma_n)$ dan iborat bo'ladi (1-rasmga qarang) va noma'lum θ parametrning bu oraliq bilan qoplanishi ehtimolligi 1-2α ga teng bo'ladi:

$P(\theta_n+z_\alpha\cdot\sigma_n<\theta<\theta_n+z_{1-\alpha}\cdot\sigma_n)=P(\theta<\theta_n+z_{1-\alpha}\cdot\sigma_n)-P(\theta<\theta_n+z_\alpha\cdot\sigma_n)=1-\alpha-\alpha=1-2\alpha.$

Endi θ_n baho asimptotik normal taqsimotga ega bo'lgan holni ko'raylik:

$$\lim_{n\to\infty}P(\frac{\sqrt{n}(\theta_n-\theta)}{\sigma_F^2}<z)=\Phi(z),$$

ya'ni

$$\sqrt{n}(\theta_n-\theta)\underset{n\to\infty}{\overset{D}{\Rightarrow}}N(0,\sigma_F^2).$$

Bu holda σ_F^2 dispersiya F ga bog'liq bo'lganligi sababli, biz yuqoridagidek aniq ishonch oralig'i qura olmaymiz. Bunday holatda odatda σ_F^2 ni biror $\hat{\sigma}_F^2$ statistika bilan baholab, θ uchun taqribiy (asimptotik) ishonch oralig'ini

$$(\theta_n-\frac{\hat{\sigma}_n\cdot t_\alpha}{\sqrt{n}};\theta_n+\frac{\hat{\sigma}_n\cdot t_\alpha}{\sqrt{n}})$$

ko'rinishida quriladi. Bu yerda $\Phi(-t_\alpha)=1-\Phi(t_\alpha)=\alpha$. Bunday oraliqlar standart oraliqlar deb ataladi. Efron bunday oraliqlarni qo'llash qoniqarsiz ekanligini takidlab o'tgani edi [1,2,4]. Buni quyidagicha tushuntirish mumkin. Biz $\sqrt{n}(\theta_n-\theta)$ ning asl $G_n(z;F)$ – aniq taqsimoti o'rniga taqribiy $\Phi(z/\hat{\sigma}_n)$ taqsimotdan foydalandik. Bunday standart oraliqlar kamchiliklarini yo'qotish uchun Efron butstrep usulini qo'llashni taklif etdi. Bu maqsadda u G_n ning butstrep bahosi bo'lgan G_m^* ning kvantillaridan foydalanishni tavsiya etdi. Buning uchun $X^{(n)}=(X_1,...,X_n)$ dastlabki tanlanmadan $Y_m^* = (X_1^*, \ldots, X_m^*)$ butstrep tanlanmani tuzib olamiz. Y_m^* дан $\theta_m^* = \theta_m^*(Y_m^*)$ bahoni θ_n formulasi asosida tuzib olib, bu jarayonni M marta takrorlab $\{\xi_m^* = \sqrt{n}(\theta^*_n - \theta^*), m=1,...M\}$ larni hosil qilamiz. $G_n(z;F)$ ning butstrep emperik analogini hisoblaymiz:

$$\hat{G}_M^*(z) = \frac{1}{M}\sum_{m=1}^{M} I(\xi_m^* < z). \qquad (1.5)$$

Endi $\hat{G}_M^*$ taqsimotning α va 1-α satxdagi ξ_α^* va $\xi_{1-\alpha}^*$ kvantillarini qo'llab quyidagi butstrep ishonch oraliqlarini

$$(\theta_n + \frac{\hat{\sigma}_n \cdot \xi^*_\alpha}{\sqrt{n}}; \theta_n + \frac{\hat{\sigma}_n \cdot \xi^*_{1-\alpha}}{\sqrt{n}})$$

yuqoridagidek 1-2α ehtimollik bilan aniqlaymiz. [1,2,4-8] maqolalarning mualliflari bunday ishonch oraliqlari asimptotik oraliqlarga nisbatan sezilarli darajada torroq ekanini ko'rsatganlar.

Biz ma'lum vaqt qulaylik uchun m=n, yani butstrep tanlanma hajmi dastlabki tanlanma hajmiga teng deb olamiz (demak (1.4) shart bajarilar ekan). Butstrep tanlanmani $Y_n^{(n)} = (X_1^*, \ldots, X_n^*)$ orqali va u yordamida tuzilgan emperik taqsimot funksiyani esa $F_n^*(x)$ deb belgilab olamiz [9]:

$$F_n^*(t) = \frac{1}{n}\sum_{k=1}^{n} I(X_k^* < t), \quad t \in R \qquad (1.6)$$

$Y_n^{(n)}$ butstrep tanlanmani tuzishning eng qulay usuli $X^{(n)} = (X_1, ..., X_n)$ tanlamadan $Y_n^{(n)} = (X_1^*, \ldots, X_n^*)$ ni takrorlanadigan tanlash usulidan iboratdir. Masalan, n=3 bo'lib, $X^{(3)} = (X_1, X_2, X_3)$ dastlabki tanlanma bo'lsin. U holda $Y_3^{(3)}$ butstrep tanlanma tarkibi $Y_3^{(3)} = (X_3, X_2, X_3)$ – ko'rinishda bo'lishi mumkin. Bu holda $X^{(3)}$ dan birinchi tanlashda $X_1^* = X_3$ ni tanlab yana qaytarib $X^{(3)}$ ga tashlaymiz, $X_2^* = X_2$ ni ham va $X_3^* = X_3$ larni ham shu tarzda tanlab, natijada yuqoridagi $Y_3^{(3)}$ ni tuzib olamiz. $Y_3^{(3)}$ dagi har bir, $\{X_k^*, k=1,2,3\}$ larni $X^{(3)}$ dan $\frac{1}{3}$ ehtimollik bilan tanlash mumkin. Demak, umumiy holda, $Y_n^{(n)}$ elementlarining har birini $X^{(n)}$ tanlanmadan $\frac{1}{n}$ ehtimolliklar bilan tanlar ekanmiz. Bu esa aynan diskret tekis taqsimot, ya'ni emperik taqsimot qonunidir.

Shunday qilib, $\{X_k^*, k = \overline{1.n}\}$ tasodifiy miqdorlarning taqsimot funksiyasi odatdagi empirik taqsimot funksiya

$$\hat{F}_n(t) = \frac{1}{n}\sum_{k=1}^{n} I(X_k < t), \; t \in R, \qquad (1.7)$$

ekan, yani (1.3) shart bajariladi. Masalan, agar $\theta = M_F X_k$ (1.1-misolga qarang) bo'lsa, u holda θ ga baho sifatida

$$\theta_n = \overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$$

deb odatdagi o'rta arifmetik qiymatni olamiz va (1.1) ga asosan

$$M_F\overline{X} = \theta, \quad D_F\overline{X} = \frac{\sigma_F^2}{n},$$

ammo

$$M_{\hat{F}_n} X^*_k = \overline{X}, \quad D_{\hat{F}_n} X^*_k = S^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X})^2$$

Demak, markaziy limit teoremaga asosan:

$$\lim_{n\to\infty} P_{\hat{F}_n}\left(\frac{\sqrt{n}(\overline{X}^* - \overline{X})}{S^*} < x\right) = \Phi(x), \qquad x \in R, \qquad (1.8)$$

bo‘lib, bu yerda $\overline{X}^{*}=\frac{1}{n}\sum_{k=1}^{n}X^{*}_{k}$, $S^{*^2}=\frac{1}{n}\sum_{k=1}^{n}(X^{*}_{k}-\overline{X}^{*})^2$. Endi θ uchun ishonch oralig‘ini qurish masalasiga to‘xtalib o‘tamiz. Biz odatdagi

$$(\overline{X}-\frac{S\cdot t_{\alpha}}{\sqrt{n}};\overline{X}+\frac{S\cdot t_{\alpha}}{\sqrt{n}})$$

ishonch oralig‘ini umumiy holda qo‘llay olmaymiz, chunki t_α soni standart normal N(0,1) taqsimotning α satxdagi kvantilidir, ammo tanlanma hajmi katta bo‘lmagan (n→∞) holda $\frac{\sqrt{n}(\overline{X}-\theta)}{S}$ ning taqsimoti (yani Styudent taqsimoti) N(0,1) dan ancha farqli bo‘lishi mumkin. Bunday hollarda yuqorida takidlab o‘tilgan Efron tavsiyasiga asosan t_α ni uning butstrep bahosi t_α^* ga almashtiramiz. t_α^* ni esa $\frac{\sqrt{n}(\overline{X}^{*}-\overline{X})}{S^{*}}$ ning taqsimoti kvantilidan yetarlicha ko‘p butstrep tanlanmalar orqali emperik usulda aniqlaymiz. Demak, θ uchun butstrep ishonch oralig‘i

$$(\overline{X}-\frac{S\cdot t^{*}_{\alpha}}{\sqrt{n}};\overline{X}+\frac{S\cdot t^{*}_{\alpha}}{\sqrt{n}})$$

bo‘ladi, ya’ni

$$\lim_{n\to\infty}\mathrm{P}(\overline{X}-\frac{S\cdot t^{*}_{\alpha}}{\sqrt{n}}<\theta<\overline{X}+\frac{S\cdot t^{*}_{\alpha}}{\sqrt{n}})=1-\alpha.$$

2.Butstrep baholar uchun limit teoremalar

Aytaylik, $X^{(n)}=(X_1,...,X_n)$ statistik tanlanma noma’lum F taqsimotga ega bo‘lgan bo‘sh to‘plamdan olingan bo‘lib, $Y_m^*=(X_1^*,\ldots,X_m^*)$ esa $X^{(n)}$ dan tuzilgan butstrep tanlanma bo‘lsin. F ga mos o‘rta qiymat va dispersiyalarni kiritamiz:

$$\theta=\theta(F)=\int_{-\infty}^{\infty}xdF(x),$$

$$\sigma^2=\int_{-\infty}^{\infty}(x-\theta)^2dF(x).$$

θ va σ^2 parametrlarni mos ravishda

$$\overline{X}_{n}=\frac{1}{n}\sum_{k=1}^{n}X_k,$$

$$S_n^{\ 2}=\frac{1}{n}\sum_{k=1}^{n}(X_k-\overline{X})^2,$$

tanlanma o‘rta arifmetik qiymat va dispersiyalar bilan baholash tabiiydir. U holda markaziy limit teoremaga asosan

$$\xi_n=\frac{\sqrt{n}(\overline{X}_n-\theta)}{S_n}\underset{n\to\infty}{\overset{D}{\Rightarrow}}\mathrm{N}(0{,}1).$$

[1] maqolada butstrep o‘rta qiymat

$$\overline{X_m^*}=\frac{1}{m}\sum_{k=1}^{m}X_k,$$

xossalari o‘rganilgan. Eslatib o‘tamiz Y_m^* butstrep tanlanma taqsimoti $\hat{F}_n$ empirik taqsimotdir ((1.7) da aniqlangan). $\overline{X_m^*}$ tanlanmaga mos butstrep tanlanma dispersiyani

$$S_m^{*^2}=\frac{1}{m}\sum_{k=1}^{m}(X_k^*-\overline{X_m^*})^2,$$

hamda ξ_n ning butstrep anologlarini

$$\xi_m^* = \frac{\sqrt{m}(\overline{X}_m^* - \overline{X}_n)}{S_n},$$

kiritamiz.

2.1-teorema [1]. $X_1, X_2, \ldots$ tasodifiy miqdorlar ketma-ketligi chekli musbat dispersiya σ^2 ga ega bo‘lsin: $0 < \sigma^2 < \infty$. U holda n, m→ ∞ da deyarli barcha tanlanmalar uchun $\sqrt{m}(\overline{X_m^*} - \overline{X_n})$ ning berilgan $X_1, X_2, \ldots, X_n$ lardagi shartli taqsimoti N(0, σ^2) ga intiladi. Bundan tashqari, shartli ehtimollik bilan $S_m^* \to \sigma$, ya’ni bir ehtimollik bilan ixtiyoriy

ε >0 учун:

$$\lim_{n,m\to\infty} \mathrm{P}(\left|S_m^* - \sigma\right| > \varepsilon / X_1, \ldots, X_n) = 0.$$

Demak, 2.1-teoremadan

$$\sqrt{m}(\overline{X_m^*} - \overline{X_n}) \underset{m\to\infty}{\overset{D}{\Rightarrow}} N(0, S_n^2) \underset{n\to\infty}{\overset{D}{\Rightarrow}} N(0, \sigma^2).$$

Bundan esa n, m→ ∞ da ξ_n va ξ_m^* tasodifiy miqdorlar ketma-ketliklarining limit taqsimoti yagona N(0,1) ekani kelib chiqadi. Bu da’vo yuqorida (1.8) munosabat bilan ta’kidlab o‘tilgan edi.

Endi oddiy empirik taqsimot funksiya (1.7) va uning butstrep analogi ((1.6) ning umumlashmasi)ni

$$F_m^*(t) = \frac{1}{m}\sum_{k=1}^{m} \mathrm{I}(X_k^* < t), \quad t \in \mathrm{R} \tag{2.1}$$

kiritamiz. Bu empirik taqsimotlar yordamida empirik jarayon

$$\omega_n(t) = \sqrt{n}(\hat{F}_n(t) - F(t))$$

va butstrep empirik jarayonlarni

$$\omega_{nm}(t) = \sqrt{m}(F_m^*(t) - \hat{F}_n(t))$$

tuzib olamiz. Matematik statistikaning noparametrik nazariyasidan ma’lumki [3], $\omega_n(t)$ tasodifiy jarayon n→∞ da “Broun ko‘prigi” deb ataluvchi Gauss jarayoni B(F(t)) ga kuchsiz yaqinlashadi:

$$\omega_n(t) \underset{n\to\infty}{\overset{D}{\Rightarrow}} \mathrm{B}(F(t)), \quad t \in \mathrm{R} \tag{2.2}$$

bu yerda $\{\mathrm{B}(y), 0 \le y \le 1\}$ o‘rta qiymati nol (MB(y)=0) va kovariyatsiyasi

$\mathrm{M\,B}(y_1)\,\mathrm{B}(y_2) = \min(y_1, y_2) - y_1 y_2$, $y_1, y_2 \in [0,1]$

formula bilan aniqlanadi. Bu (2.2) da’voning butstrep jarayon uchun analogi [1] da isbotlangandir.

2.2-teorema. [1]. Deyarli barcha tanlanma ketma-ketliklar $X_1, X_2, \ldots$ uchun $\omega_{nm}(t)$ butstrep empirik jarayonning $X_1, X_2, \ldots, X_n$ lar berilganidagi shartli taqsimotlari n, m→ ∞ da B(F(t)) ga kuchsiz yaqinlashadi:

$$\{\omega_{nm}(t) / X_1, \ldots, X_n\} \underset{\substack{n\to\infty \\ m\to\infty}}{\overset{D}{\Rightarrow}} \mathrm{B}(F(t)), \quad t \in \mathrm{R} \tag{2.3}$$

(2.3) munosabatdan hususan butstrep $F_m^*(t)$ bahoning asosliligi ham kelib chiqadi: $\forall \varepsilon > 0$ uchun

$$\lim_{n,m\to\infty} \mathrm{P}(\sup_{-\infty<t<\infty} \left|F_m^*(t) - F(t)\right| > \varepsilon / X_1, \ldots, X_n) = 0. \tag{2.4}$$

Bundan tashqari, (2.3) da’vo yordamida noma’lum taqsimot funksiya F(t) uchun asimptotik ishonch oralig‘ini tuzish mumukin. Empirik taqsimot $\hat{F}_n(t)$ va berilgan $\alpha \in (0,1)$ son uchun $C_n(\hat{F}_n)$ ni butstrep taqsimotidan shunday tanlaymizki (m=n uchun):

$$\lim_{n\to\infty} P\{\sqrt{n} \sup_{-\infty<t<\infty} \left| F_n^*(t) - \hat{F}_n(t) \right| < C_n(\hat{F}_n(t)/ X_1,...,X_n\} = 1-\alpha. \qquad (2.5)$$

munosabat bajarilsin. U holda (2.5) ga asosan

$$\lim_{n\to\infty} P\{\sqrt{n} \sup_{-\infty<t<\infty} \left| \hat{F}_n(t) - F(t) \right| \le C_n(\hat{F}_n)\} = 1-\alpha. \qquad (2.6)$$

Demak, yetarlicha katta n larda (2.6) dan 1-α ehtimollik bilan quyidagi tengsizlik bajariladi:

$$\sqrt{n}\left| \hat{F}_n(t) - F(t) \right| \le C_n(\hat{F}_n), \quad t \in R,$$

ya’ni

$$\hat{F}_n(t) - \frac{C_n(\hat{F}_n)}{\sqrt{n}} \le F(t) \le \hat{F}_n(t) + \frac{C_n(\hat{F}_n)}{\sqrt{n}}, \quad t \in R \qquad (2.7)$$

(2.7) dan noma’lum F(t) uchun 1-α satxdagi asimptotik ishonch oralig‘i quyidagicha

$$(\hat{F}_n(t) - \frac{C_n(\hat{F}_n)}{\sqrt{n}}; \hat{F}_n(t) + \frac{C_n(\hat{F}_n)}{\sqrt{n}})$$

ekan. Shuni ta’kidlab o‘tamizki,

$$C_n(\hat{F}_n) \xrightarrow[n\to\infty]{P} \beta_{1-\alpha},$$

bu yerda $\beta_{1-\alpha}$ soni $\sup\limits_{0\le y\le 1}|B(y)|$ tasodifiy miqdorning 1-α satxdagi kvantilidir

$$P(\sup_{-\infty<t<\infty} |B(F(t))| \le \beta_{1-\alpha}) = P(\sup_{0\le y\le 1}|B(y)| \le \beta_{1-\alpha}) = 1-\alpha.$$

Xulosa. Ushbu maqolada noma’lum parametrlarni statistik baholashning butstrep usuli va uning qanday hosil qilish ketma-ketligi o‘rganilgan. Butstrep usulining boshqa baholash usullaridan afzalliklari keltirilgan. Noma’lum xarakteristikalar uchun ishonch oraliqlari qurish masalalari ham tahlil qilingan. Bundan tashqari, butstrep baholari uchun limit teoremalar ham tadqiq qilingan.

Adabiyotlar

1. Bickel P.J., Freedman D.A. Some asymptotic theory for the bootstrap.// Ann. Statist., 1981, v.9, N.6, p. 1196-1217.
2. Efron B. Bootstrap methods: another look at the jackknife. // Ann. Statist., 1979, v.7, N.1, p. 1-26.
3. Боровков А.А. “Математическая статистика. Оценка параметров. Проверка гипотез”. М.: Наука, 1984, 472 с.
4. Кошевник Ю.А. Асимптотические свойства бутстреп оценок. // Заводская лаборатория 1987, т.53, N.10, с. 76-82.
5. Фармонов Ш.Қ., Абдушукуров А.А. “Математик статистика. 1-қисм: Параметрларни баҳолаш”.Т., “Университет”, 1994, 68-бет.
6. Мурадов Р.С., Зиёитдинова М.А. Бутстреп усули ва унинг статистик баҳолашда қўлланилиши. // ”Ёш математикларнинг янги теоремалари 2022” Республика конференцияси материаллари, Наманган, 13-14 май, 2022. 345-346 бетлар.
7. Michael R. Chernick Bootstrap Methods: A Guide for Practitioners and Researchers, Wiley, 2011.
8. Muradov R.S., Kamoliddinov M., Ziyoitdinova M. Taqsimotlar qorishmasini to’liq bo’lmagan tanlanmalar bo’yicha baholash //Proceedings of VI International Scientific conference STATISTICS and its applications, 19-20 October, 2022. pp. 263-264.
9. Muradov R.S., Ziyoitdinova M. Butstrep usuli yordamida statistik ishonch intervallarini qurish. NamDU Ilmiy Axborotnomasi, 11-son, 2023, 83-90 betlar.

РЕЗЮМЕ

Butstrep usulida qurilgan baholar ko‘plab yaxshi xossalarga ega bo‘ladi. Matematik statistikada ko‘pgina baholash usullari yordamida qurilgan baholarning aniq yoki limit taqsimotlari bosh to‘plamning noma’lum taqsimotiga bog‘liq bo‘lib qoladi va ularni amaliyotda qo‘llanilishi qiyinlashadi. Ushbu maqolada o‘rganilgan butstrep usulida qurilgan baholar bunday kamchiliklardan holi bo‘lib, ularni noma’lum xarakteristikalar uchun ishonch oraliqlari qurishda ham keng foydalanish mumkindir. Bundan tashqari, maqolada butstrep usulida qurilgan baholar uchun limit teoremalar ham o‘rganilgan.

РЕЗЮМЕ

Оценки, построенные с использованием метода бутстрепа, обладают многими хорошими свойствами. В математической статистике точные или предельные распределения оценок, построенные с использованием многих методов оценки, зависят от неизвестного распределения совокупности, что затрудняет их использование на практике. Оценки, построенные с помощью изучаемого в статье метода бутстрепа, лишены подобных недостатков и могут широко использоваться при построении доверительных интервалов для неизвестных характеристик. Кроме того, в статье рассматриваются предельные теоремы для оценок, построенных методом бутстрепа.

SUMMARY

Estimates constructed using the bootstrap method have many good properties. In mathematical statistics, the exact or limit distributions of estimates constructed using many estimation methods depend on the unknown distribution of the population, making them difficult to use in practice. Estimates constructed by the bootstrap method studied in this article are free of such shortcomings, and they can be widely used in the construction of confidence intervals for unknown characteristics. In addition, the article examines limit theorems for estimates constructed by the bootstrap method.