Лекция 15
Үлестірімдердің белгісіз параметрлерін бағалау

Ең алдымен эмпирикалық үлестірім функциясының әр нүктедегі мәнін осы нүктедегі теориялық үлестірім функциясының *бағасы* ретінде қарастыруға болатынын, ал әртүрлі таңдамалық сипаттамаларды (моменттер, квантилдер т.с.с) сәйкес теориялық сипаттамалардың бағалары ретінде алуға болатынын атай кетелік. Бұл жерде «баға» деген ұғымды қолдануды көлемі жеткілікті үлкен таңдамалық сипаттамалардың іске асырулары (реализациялары) мен сәйкес теориялық сипаттамалардың мәндерінің айырымының тым алшақ болуының өте аз ықтималды болатындығымен негіздеуге болады; сондықтан да (ең болмағанда көлемі өте үлкен таңдамалар үшін) таңдамалық сипаттаманы сәйкес белгісіз теориялық сипаттаманың жуық мәні ретінде алуға болады. Осылайша бұл «баға» терминіне белгілі бір асимптотикалық мағына беріледі. Бірақ та статистикалық теорияны іс жүзінде қолдану барысында қарастырылып отырған модельдің әртүрлі теориялық сипаттамалары үшін жуықтаулар құрастыруға және де бұл жуықтауларды тиімділік (қолдану тиімділігі, ыңғайлылығы, жеңілдігі т.с.с.) жағынан (неге бұл жуықтаудың ұсынылғанын қандай да бір көзқарастар тұрғысынан) негіздеуге тура келеді. Оның үстіне әдетте біздің қолымыздағы бар статистикалық деректер де шексіз көп емес, іс жүзінде олардың саны әрқашан ақырлы. Міне, осыған ұқсас есептерді шығару белгісіз параметрлерді бағалау теориясында қарастырылады.

Статистикалық бағалар және оларға қойылатын
негізгі талаптар. Бағалардың қасиеттері

Бұдан былай қарай қандай да бір ξ кездейсоқ шамасының бақыланатын мәндеріне сәйкес келетін қайталанатын тәуелсіз сынақтарға сәйкестелінген $\mathrm{F} = \{F(x;\theta),\ \theta \in \Theta\}$ параметрлік статистикалық модель бар деп есептейміз.

Айталық, $\mathrm{J}(\xi) \in \mathrm{F}$ үлестірімінен алынған $X = (X_1, X_2, ..., X_n)$ таңдамасы бар болсын. Осылайша, бақыланатын ξ кездейсоқ шамасы туралы априорлы (тәжірибеге дейінгі) біздегі бар мәлімет оның үлестірім функциясы $F(x;\theta)$ берілген F параметрлік үлестірімдер үйірінің элементі болатындығынан (яғни белгілі функционалдық формалы, бірақ белгісіз θ параметрінен тәуелді (θ берілген Θ параметрлік жиынының кез келген элементі болуы мүмкін)) тұрады.

Жалпы түрде бағалау туралы есеп былай айтылады: $X = (X_1, X_2, ..., X_n)$ *таңдамасы арқылы жеткізілген статистикалық деректерді (ақпараттарды) қолдану арқылы белгісіз* θ *параметрінің нақты (тура) бағасы* θ^0 *туралы статистикалық қорытындылар жасау керек, яғни* θ^0 *нүктесін бағалау керек.*

1-анықтама. $X = (X_1, X_2, ..., X_n)$ таңдамасының ғана функциясы болатын *кез келген кездейсоқ шама статистика* деп аталады.

Анықтамадан, мәселен, егер $T : R^n \to R$ борельдік функция болса, онда $T(X) = T(X_1, ..., X_n)$ статистика болатыны шығады, өйткені бұл жағдайда $T(X)$ – кездейсоқ шама.

Статистиканың маңызды мысалдары ретінде таңдамалық моменттерді, эмпирикалық үлестірім функциясын, реттік статистикаларды айтуға болады.

Нүктелік бағалау берілген $x = (x_1, x_2, ..., x_n)$ іске асырылымы (реализациясы) үшін θ^0 параметрінің жуық мәні ретінде қабылданатын X таңдамасының функциясы болатын $T = T(X)$ статистикасын іздейді. Бұл жағдайда $T = T(X)$ статистикасы θ параметрінің бағасы деп аталады. Әдетте (бірақ әрқашан емес) T бағасының мүмкін мәндер жиыны Θ болады.

Әрине, θ параметрін бағалау үшін әртүрлі бағаларды қолдануға болатыны түсінікті, бірақ, олардың ішіндегі белгілі бір мағыналардағы ең тиімділерін (жақсыларын) табу үшін бағалардың сапасын салыстыратын критерилер болуы қажет. Өз кезегінде, бағаларды құрастыру мақсаттарына байланысты, критерилер де әртүрлі болулары мүмкін, бірақ та кез келген критерий баға мен оның шын мәнінің жақындық өлшеміне байланысты анықталатыны түсінікті. Қарастырылатын бағалар класы әрқилы талаптармен шектелінуі қажеттігі де даусыз. Осылайша, берілген модель үшін белгілі бір бағалар класы қарастырылса және бағалардың жақындық өлшемі таңдалса, онда анықтама бойынша, жақындық өлшемін минимумдайтын баға осы кластағы оптималды (тиімді) баға болады.

Бұдан былай қарай көлемі n – ге тең $X = (X_1, X_2, ..., X_n)$ таңдамасы бойынша құрастырылған θ параметрінің бағасын көбінесе $\hat{\theta}_n = \hat{\theta}_n(X_1, ..., X_n)$ арқылы белгілейтін боламыз.

Ығыстырылмаған бағалар

Анықтама бойынша, егер кез келген $\theta \in \Theta$ үшін

$$M_\theta T(X) = \theta \qquad (1)$$

шарты орындалса, онда $T = T(X)$ статистикасы θ параметрі үшін *ығыстырылмаған* баға деп аталады.

(1)-шарт орындалмайтын бағалар үшін *ығысу функциясы* деп аталатын

$$b(\theta) = M_\theta T(X) - \theta$$

функциясын (шамасын) енгізуге болады. Онда

$$M_\theta (T-\theta)^2 = D_\theta T + b^2(\theta)$$

шамасын *қатенің орта квадраты* немесе T *бағасының орташа квадраттық қатесі* деп атайтын боламыз. Ығыстырылмаған бағалар үшін $b(\theta) = 0$, яғни орташа квадраттық қате бағаның дисперсиясымен бірдей (тепе-тең) болады.

Кейде параметр θ-ның өзін емес, оның нендей де бір $\tau(\theta)$ функциясын бағалау қажет болады ($\tau(\theta)$ функциялары *параметрлік функциялар* деп аталады). Бұл жағдайда

$$M_\theta T = \tau(\theta),\ \theta \in \Theta,$$

шартын қанағаттандыратын $T = T(X)$ сатистикасы $\tau(\theta)$ параметрлік функциясы үшін ығыстырылмаған баға деп аталады.

1-ескерту. (1)-қасиеттің маңызын бірдей үлестірімнен алынған, әрқайсысының көлемі n – ге тең N тәуелсіз таңдамалар алу мысалы негізінде түсіндіруге болады. T_i арқылы $T(X)$ бағасының i – ші таңдамадағы мәнін белгілелік: $T_i = T\left(X_1^{(i)},..., X_n^{(i)}\right)$. Егер бағалар ығыстырылмаған болса, онда $M_\theta T_i = \theta$, $T_1, T_2,..., T_n$ – тәуелсіз және бірдей үлестірілген. Ендеше күшейтілген үлкен сандар заңы бойынша

$$\frac{T_1 + T_2 + ... + T_N}{N} \xrightarrow{a.\partial.} \theta .$$

Егер $DT_i = \sigma^2$ дисперсиясы ақырлы болса, онда орталық шектік теорема бойынша

$$\frac{T_1 + T_2 + ... + T_n}{N} - \theta \overset{ac.}{\sim} \mathrm{N}\left(0, \frac{\sigma}{\sqrt{N}}\right).$$

Тиянақты, күшті тиянақты және асимптотикалық нормалды бағалар

2-анықтама. Егер $\hat\theta_n$ бағасы $\theta \in \Theta$ белгісіз параметріне $n \to \infty$ кезде *ықтималдық бойынша* жинақталса ($\hat\theta_n \xrightarrow{P_\theta} \theta$), онда $\hat\theta_n$ *тиянақты (тыңғылықты) баға* деп аталады.

Тұжырым. Егер $M_\theta \hat\theta_n = \theta$ және $D_\theta \hat\theta_n \to 0$ ($n \to \infty$) болса, онда $\hat\theta_n$ тиянақты баға.

Дәлелдеу. Шындығында да Чебышев теңсіздігі бойынша кез келген $\varepsilon > 0$ үшін

$$P_\theta\left\{\left|\hat\theta_n - M\hat\theta_n\right| > \varepsilon\right\} = P_\theta\left\{\left|\hat\theta_n - \theta\right| > \varepsilon\right\} \le \frac{D_\theta \hat\theta_n}{\varepsilon^2} \to 0 \quad (n \to \infty),$$

Ал бұл $\hat\theta_n$ бағасы θ – ға ықтималдық бойынша жинақталуының анықтамасы. ▼

Мәселен, $\hat F_n(x)$ эмпирикалық үлестірім функциясы $F(x)$ теориялық үлестірім функциясының тиянақты бағасы болатыны бұл тұжырымнан бірден шығады, себебі $M\hat F_n(x) = F(x)$, $D\hat F_n(x) = \dfrac{F(x)(1-F(x))}{n} \to 0$ ($n \to \infty$).

Оптималды бағалар

Айталық, $\mathrm{F} = \{F(x,\theta), \theta \in \Theta\}$ моделінде берілген $\tau = \tau(\theta)$ параметрлік функциясын сәйкес $X = (X_1, X_2,..., X_n)$ таңдамасы арқылы алынған (жеткізілген) ақпарат арқылы бағалау керек болсын. Берілген есепте ығыстырылмаған баға бар, яғни ығыстырылмағандық шартын қанағаттандыратын $T = T(X)$ статистикасы ($M_\theta T(X) = \tau(\theta)$ шартын қанағаттандыратын $T(X)$ статистикасы) бар болсын деп ұйғаралық. F_τ арқылы $\tau = \tau(\theta)$ параметрлік функциясының барлық ығыстырылмаған бағаларының класын белгілелік:

$$F_\tau = \{T = T(X): M_\theta T(X) = \tau(\theta)\}.$$

Қосымша, әрбір $T \in F_\tau$ үшін оның дисперсиясы ақырлы ($D_\theta T = M_\theta[T - \tau(\theta)]^2 < \infty, \quad \theta \in \Theta$) болсын деп есептелік. Бұл жағдайда бағалардың дәлдігін олардың дисперсиясы арқылы өлшеуге болады және біз F_τ класындағы әртүрлі бағаларды салыстырудың қарапайым критериін аламыз.

Айталық, T^* және T – F_τ класындағы бағалар болсын. Егер

$$D_\theta T^* \le D_\theta T \,, \; \theta \in \Theta \,, \tag{6}$$

болса, онда *дисперсияның минимумы критерии* бойынша T^* бағасы бірқалыпты (θ параметрі бойынша) түрде T бағасынан нашар емес; егер де (6)-теңсіздікте ең болмағанда бір θ үшін қатаң теңсіздік орындалса, онда T^* бағасына әлдеқайда дәлірек баға ретінде басымдық беру керек. Егер де (6) - шарт кез келген $T \in F_\tau$ үшін орындалса, онда T^* бағасы *бірқалыпты минималды дисперсиялы ығыстырылмаған баға* деп аталады. Мұндай T^* бағасын алдағы уақытта қысқаша *оптималды (тиімді) баға* деп атайтын боламыз және оны көбіне (оның $\tau(\theta)$ функциясына қатыстылығын көрсету үшін) τ^* арқылы белгілейміз.

Сонымен $\tau^* \in F_\tau$ оптималды бағасы үшін

$$D_\theta \tau^* = \inf_{T \in F_\tau} D_\theta T \,, \quad \theta \in \Theta \,. \tag{7}$$

4-ескерту. Дисперсияның бірқалыпты аз болуы туралы талап күшті талап және ол әруақытта орындала бермейді. Мәселен, $T_1, T_2 \in F_\tau$ екі бағасы үшін параметр θ – ның кейбір мәндерінде (F_τ класында) $D_\theta T_1$ минималды болуы, ал θ – ның басқадай мәндері үшін $D_\theta T_2$ минималды болуы мүмкін. Мұндай жағдайларда дисперсияның минимумының бір ғана критериі бойынша бұл бағаларды салыстыруға болмайды. Бірақ та, егер мұндай баға бар болса, бұл талап F_τ класындағы оптималды бағаны бірмәнді бөліп көрсетеді. Бұл туралы келесі теоремада айтылады.

1-теорема. Айталық, $T_1 = T_1(X)$, $T_2 = T_2(X)$ бағалары $\tau = \tau(\theta)$ параметрлік функциясы үшін F_τ класындағы екі оптималды баға болсын. Онда олар *а.д.* түрде тең:

$$T_1(X) = T_2(X) \;\; (а.д.).$$

5-ескерту. Бұдан былай қарай *статистикалардың теңдігін ақиқат дерлік түрдегі (бірге тең ықтималдықпен) теңдік* деп түсінеміз, яғни $T_1(X) = T_2(X)$ теңдігін әрқашан $P\{\omega : T_1(X) \ne T_2(X)\} = 0$ орындалған деп түсінеміз және оны $T_1(X) = T_2(X)\,(а.д.)$ түрінде жазатын боламыз. Кейде, егер мәтіннен түсінікті болса, ($а.д.$) деген сөз тіркесін жазбайтын да боламыз.

1-теореманың дәлелдеуі. Жаңа $T_3 = \dfrac{T_1 + T_2}{2}$ бағасын қарас-тыралық. Әрине, $T_3 \in F_\tau$ болатыны түсінікті және де

$$D_\theta T_3 = \frac{D_\theta T_1 + D_\theta T_2 + 2\operatorname{cov}_\theta(T_1, T_2)}{4}. \tag{8}$$

Мынандай белгілеулер енгізелік: $u = D_\theta T_i \;\; (i = 1,2)$. Онда $D_\theta T_3 \ge u$ (өйткені T_1 және T_2 оптималды бағалар), ал Коши-Буняковский теңсіздігі бойынша $\left|\operatorname{cov}_\theta(T_1, T_2)\right| \le \sqrt{D_\theta T_1} \cdot \sqrt{D_\theta T_2} = u$. Демек (8)-ден мынаны аламыз:

$$D_\theta T_3 \le \frac{u + u + 2u}{4} = u.$$

Ары қарай, $D_\theta T_3 \ge u$ теңсіздігін еске алып, $D_\theta T_3 = u = D_\theta T_i \;\; (i = 1,2)$ теңдігін аламыз, яғни T_3 – те оптималды баға. Сонымен

$$D_\theta T_3 = \frac{u + \operatorname{cov}_\theta(T_1, T_2)}{2} = u \,,$$

бұдан $\operatorname{cov}_\theta(T_1, T_2) = D_\theta T_1 = u \ge 0$ болатынын аламыз. Демек T_1 мен T_2 *а.д.* түрде сызықтық байланысты (өйткені корреляция коэффициенті $\rho_{T_1, T_2} = 1$), яғни

$$T_1 = kT_2 + c \quad (а.д.)$$

Бағалардың ығыстырылмағандық шартынан

$$\tau = M_\theta T_1 = kM_\theta T_2 + c = k\tau + c,$$
$$c = (1 - k)\tau,$$

ал бұдан

$$T_1 = \tau + k(T_2 - \tau) \quad (a.\partial.)$$

теңдігін аламыз. Мұндағы k θ – параметрінің функциясы $(k = k(\theta))$ және ол келесідей теңдіктер тізбесінен анықталады:

$$u = \operatorname{cov}_\theta(T_1, T_2) = M_\theta(T_1 - \tau)(T_2 - \tau) = M_\theta(T_2 - \tau)^2 = kD_\theta T_2 = ku\ .$$

Соңғы теңдіктен $k \equiv 1$ болатынын аламыз, демек $T_1 \equiv T_2$ $(a.\partial.)$.

Шындыққа сәйкестік функциясы.

5-анықтама. Айталық, $X = (X_1, X_2, ..., X_n)$ – таңдама, ал $f_{X_i}(x) = f_{X_i}(x,\theta)$ - X_i – дің үлестірім тығыздығы (немесе $f_{X_i}(x) = P_\theta\{X_i = x\}$ ықтималдығы) болсын. Онда

$$L(x;\theta) = L(x_1,..., x_n;\theta) = f_{X_1}(x_1;\theta) f_{X_2}(x_2;\theta)... f_{X_n}(x_n;\theta)$$

функциясы ($\theta \in \Theta$ параметрінің функциясы ретінде) *шындыққа сәйкестік функциясы* деп аталады. Ал

$$U(X;\theta) = \frac{\partial \ln L(X;\theta)}{\partial\theta} = \sum_{i=1}^{n} \frac{\partial \ln f(X_i;\theta)}{\partial\theta}$$

функциясы X таңдамасының *үлесі* (немесе *үлес функциясы*) деп аталады.

$$i_n(\theta) = D_\theta U(X;\theta) = M_\theta U^2(X;\theta)$$

функциясы θ параметрі жөнінде X таңдамасында болатын *Фишер ақпараты функциясы* немесе *Фишер ақпараты* деп аталады.

Максималды шындыққа сәйкестік бағасының анықтамасы және мысалдары

Максималды шындыққа сәйкестік әдісі белгісіз параметрлердің бағасын табудың ең жан-жақты (универсал) әдістерінің бірі болып табылады. Белгісіз θ параметрінің бұл әдіс бойынша табылатын бағасын $\hat\theta = \hat\theta(X)$ арқылы, ал параметрлік $\tau(\theta)$ функциясының бағасын $\hat\tau = \hat\tau(\theta)$ арқылы белгілейтін боламыз.

Айталық, $X = (X_1, X_2, ..., X_n)$ таңдамасы үлестірім функциясы $F_{X_i}(x) \in \mathcal{F} = \{F(x,\theta),\ \theta \in \Theta\}$ болатын таңдама және E бұл таңдаманың $x = (x_1, x_2, ..., x_n)$ нәтижесі (іске асырылуы) үшін шындыққа сәйкестік функциасы $L(x;\theta) = L(x_1, x_2, ..., x_n;\theta)$ болсын.

11-анықтама. θ параметірінің $\hat\theta$ максималды (ең үлкен) шындыққа сәйкестік бағасы (дәлірек айтсақ, X таңдамасының берілген x іске асырылуындағы максималды шындыққа сәйкестік бағасының мәні деп берілген x үшін $L(x,\theta)$ шындыққа сәйкестік функциясы максималды мәнін қабылдайтын параметрлік Θ жиынының $\hat\theta$ нүктесін айтамыз. Сонымен $L(x;\hat\theta) \geq L(x;\theta), \theta \in \Theta$ немесе $L(x;\hat\theta) = \sup\limits_{\theta\in\Theta} L(x;\theta)$.