Лекция 14
Таңдамалық сипаттамалар. Таңдамалық моменттер

Айталық, $X = (X_1, X_2, ..., X_n) - G(\xi)$ үлестірімінен алынған таңдама, ал $F(x)$ және $\hat{F}_n(x)$ – сәйкес теориялық және эмпи-рикалық үлестірім функциялары болсын. Алдыңғы параграфта біз $F(x)$ функциясына $\hat{F}_n(x)$ функциясын сәйкес қойғанымыз секілді, $F(x)$ арқылы анықталған кез келген $g = \int g(x) dF(x)$ сипаттамасына оның статистикалық “көшірмесі”

$$G = \int g(x) d\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} g(X_i)$$

шамасын сәйкес қоюға болады. G кездейсоқ шамасын теориялық сипаттама g –ға сәйкес *эмпирикалық* немесе *таңда-малық сипаттама* деп атайды. Осылайша, таңдамалық сипат-тама – ол $X = (X_1, X_2, ..., X_n)$ таңдамасының элементтері үшін $g(x)$ функциясының арифметикалық ортасы.

Егер $g(x) = x^k$ болса, онда $G - k$ - *ші ретті таңдамалық момент* (оны a_{nk} арқылы белгілейік):

$$a_{nk} = a_{nk}(X) = \frac{1}{n}\sum_{i=1}^{n} X_i^k . \qquad (1)$$

Егер $k = 1$ болса, онда a_{n1} *таңдамалық орта* деп аталады да $\overline{X}$ арқылы белгіленеді.

Сол сияқты k –ші ретті *таңдамалық орталық момент*

$$b_{nk} = b_{nk}(X) = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^k \qquad (2)$$

қатынасы арқылы анықталады. Екінші ретті таңдамалық орталық момент b_{n2} *таңдамалық дисперсия* деп аталады да, $S^2 = S^2(X)$ арқылы белгіленеді:

$$S^2 = b_{n2}(X) = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 . \qquad (2')$$

Осы жерде және бұдан былай қарай үнемі теориялық k –ші моментті α_k, теориялық k -ші ретті орталық моментті μ_k ар-қылы белгілейтін боламыз:

$$\alpha_k = M\xi^k, \quad \beta_k = M(\xi - \alpha_1)^k. \qquad (3)$$

Таңдамалық және таңдамалық орталық моменттер α_k және μ_k өздерінің теориялық моменттері секілді байланысатынын байқау қиын емес. Жалпы жағдайда $k \ge 2$ үшін (2) қатынастан

$$b_{nk} = \sum_{r=0}^{k} (-1)^r C_k^r \left(\overline{X}\right)^r a_{n,k-r} =$$

$$= \sum_{r=0}^{k-2} (-1)^r C_k^r \left(\overline{X}\right)^2 a_{n,k-r} + (-1)^{k-1}(k-1)\left(\overline{X}\right)^k.$$

Жеке жағдайларда

$$\begin{aligned} &S^2 = a_{n2} - \left(\overline{X}\right)^2, \; b_{n3} = a_{n3} - 3\overline{X}a_{n2} + 2\left(\overline{X}\right)^3, \\ &b_{n4} = a_{n4} - 4\overline{X}a_{n3} + 6\left(\overline{X}\right)^2 a_{n2} - 3\left(\overline{X}\right)^4. \end{aligned} \qquad (4)$$

Таңдамалық абсолютті моменттер, таңдамалық семиинва-рианталар т.с.с. ұғымдар ұқсас түрде енгізіледі. Мәселен, таңдамалық абсолютті k –ші момент $\frac{1}{n}\sum_{i=1}^{n}\left|X_i\right|^k$, k -ші ретті абсолютті таңдамалық орталық момент $\frac{1}{n}\sum_{i=1}^{n}\left|X_i - \overline{X}\right|^k$ т.с.с.

Таңдамалық орта мен таңдамалық дисперсияның моменттері

$X = (X_1, X_2, ..., X_n)$ таңдамасының компоненталары тәуелсіз және бақыланатын ξ кездейсоқ шамасымен бірдей үлестірілгендіктен (төменде кейде $M\xi = a,\; D\xi = \sigma^2$ белгілеу-лерін пайдаланатын боламыз)

$$M\overline{X} = \frac{1}{n}\sum_{i=1}^{n} MX_i = M\xi = \alpha_1 = a, \; D\overline{X} =$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} DX_i = \frac{1}{n}D\xi = \frac{\mu_2}{n} = \frac{\sigma^2}{n}. \qquad (5)$$

S^2 таңдамалық дисперсияның моменттерін табалық. Алдымен кез келген c үшін

$$X_i - \overline{X} = X_i - c - \left(\overline{X} - c\right) = X_i - c - \left(\frac{1}{n}\sum_{j=1}^{n} X_j - c\right) =$$

$$= X_i - c - \frac{1}{n}\sum_{j=1}^{n}\left(X_j - c\right)$$

болатынын, яғни b_{nk} k –ші ретті таңдамалық орталық моменттер ξ кездейсоқ шамасының мәндерінің шкаласында санаудың басына (басталу нүктесіне) тәуелсіздігін байқалық. Бұл бізге басынан бастап-ақ $\alpha_1 = M\xi = 0$ деп ұйғаруға мүм-кіндік береді (басқа жағдайларда жаңа $\tilde{\xi} = \xi - \alpha_1$ кездейсоқ шамаларына көшеміз). Осыны ескерсек, онда (3), (4) қатынастардан

$$MS^2 = Ma_{n2} - M\left(\overline{X}\right)^2 = \mu_2 - \mu_2 / n = \frac{n-1}{n}\mu_2 = \frac{n-1}{n}\sigma^2. \qquad (6)$$

Ары қарай (3) формуланы пайдаланып былай жаза аламыз:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i^2 - 2X_i\overline{X} + \left(\overline{X}\right)^2\right) = \frac{1}{n}\sum_{i=1}^{n}X_i^2 - \left(\overline{X}\right)^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n}X_i^2 - \frac{1}{n^2}\sum_{i,j=1}^{n}X_iX_j = \frac{n-1}{n^2}\sum_{i=1}^{n}X_i^2 - \frac{2}{n^2}\sum_{i<j}X_iX_j .$$

Бұдан

$$\left(S^2\right)^2 = \frac{(n-1)^2}{n^4}\sum_{i=1}^{n}X_i^4 + \frac{2(n-1)^2}{n^4}\sum_{i<j}X_i^2X_j^2 - \frac{4(n-1)}{n^4}\sum_{i=1}^{n}X_i^2 \times$$

$$\times \sum_{i<j}X_iX_j + \frac{4}{n^4}\sum_{i<j}X_i^2X_j^2 + \frac{4}{n^4}\sum_{\substack{i<j,l<k \\ i\neq l \vee j\neq k}}X_iX_jX_lX_k .$$

Егер $i = j$ үшін $MX_i = M\xi = 0$ екенін және $i \neq j$ үшін X_i мен X_j–дің тәуелсіздігін ескерсек, онда $M\left(S^2\right)^2$ есептеген кезде үшінші қосынды мен соңғы бесінші қосындының математикалық күтімдері нөлге тең болатынын көреміз. Демек

$$M\left(S^2\right)^2 = M\left[\frac{(n-1)^2}{n^4}\sum_{i=1}^{n}X_i^4 + \frac{2(n-1)^2+4}{n^4}\sum_{i<j}X_i^2X_j^2\right] =$$

$$= \frac{(n-1)^2}{n^3}n\mu_4 + \frac{2(n-1)^2+4}{n^4}\cdot\frac{n(n-1)}{2}\mu_2^2 = \frac{(n-1)^2}{n^3}\mu_4 +$$

$$+ \frac{(n-1)^2+2}{n}\cdot(n-1)\mu_2^2 .$$

Ары қарай осы алынған қатынастан және (6) формуладан S^2 –тың дисперсиясы үшін

$$DS^2 = M\left(S^2\right)^2 - \left(MS^2\right)^2 = \frac{(n-1)^2}{n^3}\left[\mu_4 - \frac{n-3}{n-1}\mu_2^2\right] \qquad (7)$$

формуласын аламыз.

Жалпы, кез келген k үшін $n \to \infty$ кезде (төмендегі екінші формула $\mu_{2mk} < \infty$ болғанда дұрыс)

$$Mb_{nk} = \mu_k + O\left(\frac{1}{n}\right),\ M\left(b_{nk} - \mu_k\right)^{2m} = O\left(\frac{1}{n^m}\right)$$

болатынын көрсетуге болады ([16], 27.5).

Таңдамалық моменттердің асимптотикалық түрлері

Енді таңдаманың көлемі n өскен кезде таңдамалық моменттердің асимптотикалық түрлері қандай болатынын қарастыралық. Бізде

$$Ma_{nk} = \frac{1}{n}\sum_{i=1}^{n} MX_i^k = M\xi^k = \alpha_k,$$

$$Da_{nk} = \frac{1}{n^2}\sum_{i=1}^{n} DX_i^k = \frac{1}{n}D\xi^k = \frac{1}{n}\left[M\xi^{2k} - \left(M\xi^k\right)^2\right] = \frac{\alpha_{2k} - \alpha_k^2}{n}. \quad (8)$$

Егер $\alpha_{2k} < \infty$ болсын деп есептесек, онда Чебышев теңсіздігі бойынша кез келген $\varepsilon > 0$ үшін

$$P\left\{\left|a_{nk} - \alpha_k\right| > \varepsilon\right\} \le \frac{Da_{nk}}{\varepsilon^2} = \frac{\alpha_{2k} - \alpha_k^2}{n\varepsilon^2} \to 0\ (n \to \infty).$$

Басқаша айтқанда, егер α_{2k} теориялық моменті ақырлы болса, онда таңдаманың көлемі $n \to \infty$ кезде таңдамалық мо-менттер сәйкес теориялық моменттерге ықтималдық бойынша жинақталады:

$$a_{nk} = a_{nk}(X) \xrightarrow{P} \alpha_k\ \ (n \to \infty). \qquad (9)$$

Ұқсас тұжырым таңдамалық орталық моменттер үшін де дұрыс $\left(b_{nk} \xrightarrow{P} \beta_k, n \to \infty\right)$, жалпы ақырлы санды a_{nk} таңдамалық моменттерінің үзіліссіз функциясы болатын кез келген таңдамалық сипаттамалар үшін де дұрыс болады. Бұл тұжырым ықтималдық бойынша жинақталудың төменде 1-теорема ретінде тұжырымдалған қасиетінің ([19], 277-бет, 2-тұжырым) салдары.

1-теорема. Айталық , $\eta_{1n}, \eta_{2n}, ..., \eta_{mn}$ кездейсоқ шамалары $n \to \infty$ кезде сәйкес $\eta_1, \eta_2, ..., \eta_m$ кездейсоқ шамаларына ықти-малдық бойынша жинақталсын: $\eta_{in} \xrightarrow{P} \eta_i$, $n \to \infty$. Онда кез келген үзіліссіз $g(x_1, x_2, ..., x_k)$ функциясы үшін

$$g(\eta_{1n}, \eta_{2n}, ..., \eta_{mn}) \xrightarrow{P} g(\eta_1, \eta_2, ..., \eta_m) \ .$$

Дәлелдеу. Теореманы $g(x)$ функциясы бір айнымалының функциясы болатын жағдайдағыдай әдіспен ([19], 277-бет) дәлелдеуге болады. Шындығында да, $g(x_1, ..., x_n)$ үзіліссіз бол-ғандықтан кез келген $\varepsilon > 0$ үшін $\delta > 0$ табылады да $\{|\eta_{1n} - \eta_1| < \delta, ..., |\eta_{mn} - \eta_m| < \delta\} \subseteq \{|g(\eta_{1n}, ..., \eta_{mn}) - g(\eta_1, ..., g_m)| < \varepsilon\}$ ендіруі орындалады. Бұдан

$$\{|g(\eta_{1n}, ..., \eta_{mn}) - g(\eta_1, ..., \eta_m)| \geq \varepsilon\} \subseteq \bigcup_{i=1}^{m} \{|\eta_{in} - \eta_i| \geq \delta\},$$

$P\{|g(\eta_{1n}, ..., \eta_{mn}) - g(\eta_1, ..., \eta_m)| \geq \varepsilon\} \leq \sum_{i=1}^{m} P\{|\eta_{in} - \eta_i| \geq \delta\} \to 0 \ (n \to \infty)$, себебі шарт бойынша, $\eta_{in} \xrightarrow{P} \eta_i$. ▼

Енді жоғарыда алынған нәтижелерді жинақтап, теорема ретінде тұжырымдалық.

2-теорема. $X = (X_1, X_2, ..., X_n) - G(\xi)$ үлестірімінен алын-ған таңдама және бақыланатын кездейсоқ шама ξ –дің төменде есептеулер барысында кездесетін барлық моменттері бар және ақырлы болсын. Онда таңдамалық және таңдамалық орталық моменттер сәйкес теориялық моменттерге және теориялық орталық моменттерге ықтималдық бойынша жинақталады:

$$a_{nk}(X) = \frac{1}{n} \sum_{i=1}^{n} X_i^k \xrightarrow{P} \alpha_k = M\xi^k,$$

$$b_{nk}(X) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^k \xrightarrow{P} \beta_k = M(\xi - \alpha_1)^k. \qquad (10)$$

Тіптен кез келген үзіліссіз $g(x_1, x_2, ..., x_k)$ функциясы үшін

$$g(a_{n1}, a_{n2}, ..., a_{nk}) \xrightarrow{P} g(\alpha_1, \alpha_2, ..., \alpha_k). \qquad (11)$$

Бұл теореманың тағы бір статистикалық қолданымы туралы айта кетелік. Үзіліссіз кездейсоқ шаманың тығыздығының қасиеттерін зерттеген кезде көбінесе *асимметрия (симметриялы еместік) коэффициенті* (γ_1) және *эксцесс* (γ_2) деп аталатын және

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}}, \ \gamma_2 = \frac{\mu_4}{\mu_2^2} - 3 \qquad (12)$$

формулалары арқылы анықталатын сипаттамаларын қарастырады. Егер үлестірім тығыздығының графигі асимметриялы болса, онда $\gamma_1 = 0$ (мәселен, $\xi \sim N\left(a, \sigma^2\right)$ нор-маль кездейсоқ шама үшін $\gamma_1 = M\left(\xi - a\right)^3 = 0$), сондықтан γ_1–дің мәні арқылы тығыздықтың симметриядан ауытқу дәрежесін бағалайды. Сол сияқты нормаль үлестірім үшін $\gamma_2 = 0$, сон-дықтан эксцессі нөлге тең болатын үлестірім тығыздықтарын *нормаль эксцесті тығыздықтар* деп атайды. Егер $\gamma_2 > 0$ не $\gamma_2 < 0$ болса, онда тығыздық сәйкес *оң эксцесті* немесе *теріс эксцесті тығыздық* деп аталады.

Егер $X = \left(X_1, X_2, ..., X_n\right)$ таңдамасы абсолютті үзіліссіз $G(\xi)$ үлестірімінен алынса, онда (12)-формулаларға сәйкес *таңдамалық асимметрия коэффициенті* Γ_{n1} мен *таңдамалық эксцесс коэффициенті* Γ_{n2} мына

$$\Gamma_{n1} = \frac{b_{n3}}{b_{n2}^{3/2}} = \frac{b_{n3}}{S_n^3}, \ \Gamma_{n2} = \frac{b_{n4}}{b_{n2}^2} - 3 = \frac{b_{n4}}{S_n^4} - 3 \qquad (12')$$

формулалары арқылы анықталады (жоғарыда S_n^2 арқылы біз (2') формуласымен анықталған S^2 таңдамалық дисперсиясын бел-гіледік).

Анықтамадан көріп тұрғанымыздай Γ_{n1} мен Γ_{n2} ($\mu_2 > 0$ болған кезде) a_{nk} таңдамалық моменттерінің үзіліссіз функ-циясы, демек 2-теоремаға сәйкес, $n \to \infty$ кезде олар өздерінің теориялық сипаттамаларына ықтималдық бойынша жинақ-талады:

$$\Gamma_{n1} \xrightarrow{P} \gamma_1, \quad \Gamma_{n2} \xrightarrow{P} \gamma_2 . \qquad (13)$$

Енді таңдамалық моменттердің асимптотикалық үлестірім заңдары қандай болатынына тоқталалық .

3-теорема. Егер α_{2k} теориялық моменті ақырлы болса, онда таңдаманың көлемі $n \to \infty$ кезде

$$a_{nk} = a_{nk}\left(X_1, ..., X_n\right) \overset{ас.}{\sim} N\left(\alpha_k, \frac{\alpha_{2k} - \alpha_k^2}{n}\right), \qquad (14)$$

басқаша айтқанда

$$\tilde{a}_{nk}(X)=\frac{a_{nk}(X)-\alpha_k}{\sqrt{\frac{\alpha_{2k}-\alpha_k^2}{n}}}=\frac{\sqrt{n}\left(a_{nk}(X)-\alpha_k\right)}{\sqrt{\alpha_{2k}-\alpha_k^2}}\overset{ac}{\sim}N\ (0,1),\qquad (14')$$

яғни үлестірім функциялар тізбегі $n\to\infty$ кезде

$$F_{\tilde{a}_{nk}}(x)\to\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{u^2}{2}}du\qquad (14')$$

жинақталуын қанағаттандыратын тізбек.

Дәлелдеу. Алдымен теоремада келтірілген тұжырымдарға байланысты анықтамаларды еске түсіре кетелік (V тарау, §1; [19], IV тарау, §2, п. 2.2.).

Егер η_n кездейсоқ шамалар тізбегі үшін қандай да бір $A_n, B_n>0$ сандық тізбектері табылып, $\tilde{\eta}_n=\frac{\eta_n-A_n}{B_n}$ кездейсоқ шамалар тізбегі үшін $n\to\infty$ кезде

$$F_{\tilde{\eta}_n}(x)=P\{\tilde{\eta}_n\le x\}\to\Phi(x)$$

жинақталуы орындалса, онда біз $\tilde{\eta}_n=\frac{\eta_n-A_n}{B_n}$ кездейсоқ шамалар тізбегі стандартты нормаль кездейсоқ шамаға әлсіз жинақталады немесе бұл тізбек асимптотикалық нормаль тізбек дейтінбіз және оны қысқаша $\tilde{\eta}_n\xrightarrow{әлсіз}\eta\sim N\ (0,1)$ немесе $\tilde{\eta}_n\overset{ac.}{\sim}N\ (0,1)$ түрінде, ал бастапқы η_n тізбегі үшін бұл тұжы-рымды $\eta_n\overset{ac.}{\sim}N\ \left(A_n, B_n^2\right)$ түрінде жазатынбыз.

Сонымен теореманың тұжырымы таңдаманың көлемі өскен $(n\to\infty)$ кезде таңдамалық моменттерді нормаль кездейсоқ шамалармен жуықтауға болатынын (оның параметрлері (14)-қатынаста көрсетілген) білдіреді.

Дәлелдеуі. $a_{nk}(X)$–тәуелсіз бірдей үлестірілген кездейсоқ шамалардың қосындысы. Ендеше бұл тізбекке орталық шектік теореманы (V тарау, §1, п.1.1.2.) қолдануға болады. Бұл теоремаға сәйкес

$$\frac{a_{nk}(X)-Ma_{nk}(X)}{\sqrt{Da_{nk}(X)}}=\frac{a_{nk}-\alpha_k}{\sqrt{\frac{\alpha_{2k}-\alpha_k^2}{n}}}\overset{ac.}{\sim}N\ (0,1),$$

дәлелдеу керегі де осы еді.

Дәлелденген теореманы таңдамалық моменттер мен сәйкес теориялық моменттердің берілген ауытқуының ықтималдығын табу үшін қолдануға болады. Шындығында да кез келген $t > 0$ үшін $n \to \infty$ кезде

$$P\left\{ \frac{\sqrt{n}\left|a_{nk}(X) - \alpha_k\right|}{\sqrt{\alpha_{2k} - \alpha_k^2}} < t \right\} \approx \frac{1}{\sqrt{2\pi}} \int_{-t}^{t} e^{-\frac{x^2}{2}} dx = 2\Phi_0(t) = 2\Phi(t) - 1 ,$$

мұндағы $\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{x^2}{2}} dx$. Егер $t = 0.01$ болса, онда

$$P\left\{ \left|a_{nk}(X) - \alpha_k\right| < t^* \right\} \approx 2\Phi_0(0.01) = 0.008 ,$$

мұндағы $t^* = 0.01 \cdot \sqrt{\frac{\alpha_{2k} - \alpha_k^2}{n}}$, ал α_k, α_{2k} шамалары таңдама арқылы анықталады, n -таңдаманың көлемі.

Ары қарай көп өлшемді орталық шектік теореманы пайдаланып кез келген ақырлы санды a_{nk} таңдамалық моменттерінің бірлескен үлестірім заңының асимптотикалық нормалдылығын дәлелдеуге болатынын айта кетелік. Бұл жағдайда тек

$$Ma_{nk} = \alpha_k, Ma_{nk}^2 = \alpha_{2k}, Da_{nk} = \alpha_{2k}^2 - \alpha_k^2 ,$$

$$\operatorname{cov}(a_{nk}, a_{ns}) = Ma_{nk}a_{ns} - Ma_{nk}Ma_{ns} = \frac{1}{n^2}\sum_{i,j=1}^{n} M\left(X_i^k X_j^s\right) -$$

$$-\alpha_k\alpha_s = \frac{1}{n^2}\sum_{i=1}^{n} MX_i^{k+s} + \frac{1}{n^2}\sum_{i<j}\left(MX_i^k \cdot MX_j^s\right) - \alpha_k\alpha_s =$$

$$= \frac{\alpha_{k+s}}{n} + \frac{n-1}{n}\alpha_k\alpha_s - \alpha_k\alpha_s = \frac{\alpha_{k+s} - \alpha_k\alpha_s}{n}$$

болатынына назар аударамыз.

Орталық таңдамалық моменттер үшін де олардың асимпто-тикалық нормалдылығын дәлелдеуге болады. Мәселен, екінші ретті орталық таңдамалық момент (таңдамалық дисперсия) үшін мына теорема дұрыс: