

Лекция 2. Биомедицинадағы математикалық әдістер. Сандық таксономия негіздері. (дәріс беруші-қауымдастырылған профессор Маусымбекова С.Д.)

Алматы, 2024

Сандық Таксономия Биологияның басты проблемаларының бірі - кез-келген ағзаға ат қойғаннан кейін оны тиісті жүйелі топқа орналастыру. Жүйелілік организмдердің әртүрлілігін ғылыми тұрғыдан зерттеуді, соның ішінде осы организмдердің сипаттамасын және белгілі бір топтарға орналасуын ғана емес, сонымен бірге осы келісімділікті түсіндіруді қамтиды. Жіктеу - белгілі бір критерийлерге негізделген организмдердің топтарға орналасуы, мысалы, морфологиялық ұқсастық немесе жалпы шығу тегіне байланысты жіктеу. Таксономия теориясы, ережелері, әдістері мен қосымшалары бар жіктеу ғылымы ретінде анықталады. Таксон белгілі жіктеу әдісін қолдану арқылы алынған таксономикалық топ болып табылады. Организмдерді бір таксонға 1) елеулі фенетикалық жақындығы; 2) ортақ шығу тегі туралы белгілері арқылы немесе 3) нақты туыстық белгілері арқылы жатқызуға болады. Топтар монотетикалық және политетикалық, яғни бірдей немесе әртүрлі сипаттамалары бар топтар болып бөлінеді.

Сандық таксономия негіздері Таксономикалық зерттеуде қолданылатын негізгі бірлік операбельдік таксономикалық бірлік немесе ОТБ деп аталады. Төменде сипатталған ұқсастықты есептеудің әртүрлі әдістері кез-келген таксономикалық бірліктерге қолданылады, яғни олар жеке организмдер немесе бірліктердің абстракты класы болуы мүмкін. Монотетикалық және политетикалық топтардан өзге жіктелудің қарапайым түрі қандай да бір ерекшеліктің бар болуы немесе жоқ болуына сәйкес дихотомдық жіктеу болып табылады; бұл жағдайда ерекшеліктің бар болуы 1-ге теңелсе, ал болмауы 0-ге теңеледі. Таксондардың математикалық әдістер арқылы құрылу процедурасының негізгі кезеңдері:

1-кезең. Таңбалау үшін қолданылатын t таксономикалық бірліктерді таңдап, әрқайсысы үшін тиісті сипаттамалардың жеткілікті n мөлшерін зерттеп, оларға код беріледі. Әдетте, $n = 50-100$ немесе одан көп болуы мүмкін. Егер n тым аз болса, онда таксондардың соңғы құрылымы қосымша сипаттамаларға өте сезімтал болады және ұқсастық коэффициенттерінің статистикалық дәлдігі тым аз болады. Бастапқы деректерді $n \times t$ матрицасы түрінде жазуға болады.

2-кезең. Ұқсастық дәрежесін анықтау мақсатында сандық кодтары бар ерекшеліктердің негізінде барлық ОТБ-терге бір-бірімен салыстыру жүргізіледі. Өлшенген ұқсастық коэффициенттерінің жиынтығы $t \times t$ матрицасы ретінде беріледі. Бұл матрица симметриялық болып табылады, сондықтан негізгі диагональ бойынша жоғары немесе одан төмен жатқан элементтердің жиыны келесі зерттеулерге жеткілікті.

3-кезең. Есептелген ұқсастық коэффициенттерінің матрицасы негізінде барлық ОТБ-ді топтар бойынша біріктіруге болады. Кез-келген топта таксономикалық бірліктер әртүрлі топтарға жататын бірліктерге қарағанда әлдеқайда көп ұқсастықтарға ие. Көптеген жағдайларда бірнеше топты жоғарғы дәрежедегі топтарға біріктіру арқылы иерархиялық құрылымды алуға болады. Осындай таксономикалық схеманы кейін дендрограмма түрінде ұсынуға болады.

4-кезең. Жіктеудің дұрыстығы эксперименттік бағалаулар арқылы жүзеге асырылады.

Осылайша, жоғарыда нақты логикалық негізі бар таксономикалық талдау әдісі ұсынылған. Ассоциация коэффициенттері, корреляция коэффициенттері және қашықтық индикаторлары – ұқсастықты сандық бағалаудың үш негізгі жолы.

Ұқсастықты бағалау. Ассоциация коэффициенті Биология, экология, зоология, психология және т.б. ғылым салаларында ассоциация коэффициенттері қолданылады. Ұқсастықты бағалаудың аталған түрі қолданылатын таксономикалық бірліктердің белгілі жұбына арналған екі сипаттамалардың жиынтығын салыстыру негізінде құрылады.

Мысалы, j және k таксономикалық бірліктерін қарастырайық. Екі ОТБ – де бар сипаттаманың саны n_{11} , ал жоқ сипаттаманың саны n_{00} тең болсын; бірінші таксономикалық бірлікте бар, бірақ екінші таксономикалық бірлікте жоқ сипаттаманың саны n_{10} , және кері комбинацияны n_{01} арқылы белгіленсін. Сондай-ақ, қайталанатын m және қайталанбайтын ерекшеліктердің жұптарының санын u арқылы жазуға болады $m = n_{00} + n_{11}$, $u = n_{01} + n_{10}$, $m + u = n$. Осылайша, екі ОТБ-де бар белгілер саны m , ал u сәйкестік жоқ белгілер саны.

2-Кесте. Екі таксономикалық бірліктің ерекшеліктерін салыстыру

		0	1	Қосынды
к-шы О.Т.Б	0	n_{00}	n_{01}	n_{0*}
j-шы О.Т.Б	1	n_{10}	n_{11}	n_{*0}
	Сумма	n_{*0}	n_{*1}	n

Ең қарапайым ассоциациялық коэффициент - сәйкес келетін белгілердің қатынасы:

$$S_1 = \frac{m}{n}.$$

Бұл қатынас келесі қасиетке ие: $m \rightarrow n$ жағдайында $S_1 \rightarrow 1$, ($u \rightarrow 0$); $m \rightarrow 0$ жағдайында $S_1 \rightarrow 0$. Осылайша, толық сәйкестік бір санымен анықталса, толық сәйкессіздік нөл санымен анықталады.

Белгілер жоқ жағдайындағы сәйкестікті ескермеген жағдайда ұқсастық коэффициентін келесідей есептеуге болады:

$$S_2 = \frac{n_{11}}{n_{11} + u},$$

толық сәйкестік жағдайында $u \rightarrow 0$ және $S_2 \rightarrow 1$, ал толық сәйкессіздік жағдайында $n_{11} \rightarrow 0$, $S_2 \rightarrow 0$.

Корреляция коэффициенті. Қашықтық көрсеткіштері Корреляция коэффициенттері екі сипаттаманың арасындағы байланыс дәрежесін анықтау үшін пайдаланылады және олар үзіліссіз айнымалыларға қолданылады. Мысалы, негізгі матрица $n \times t$ барлық элементтері үзіліссіз айнымалы болып табылсын. Бұл жағдайда j - таксономикалық бірлік үшін x_{ij} , $i = 1, \dots, n$ арқылы белгіленген n өлшемдер бар. Осы шамалардың орташа мәні (\bar{x}_j) болсын. Сол сияқты, k -таксономикалық бірлік үшін x_{ik} , $i = 1, \dots, n$ (\bar{x}_k

- орташа мәні) өлшемдері бар. Бұл екі өлшемдер топтары арасындағы аралас сәттердің корреляциялық коэффициентті келесідей анықтауға болады:

$$r_{jk} = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)}{[\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \sum_{i=1}^n (x_{ik} - \bar{x}_k)^2]^{1/2}}.$$

Бұл коэффициент "-1"-ден (толық теріс корреляция жағдайы деп саналады) "+1"-ге дейін (толық оң корреляция жағдайы) өзгереді. Екі таксономикалық бірлік арасындағы ұқсастық дәрежесін анықтаудың тағы бір жолы - олардың арасындағы «қашықтық» көрсеткішін пайдалану. Мысалы, тек үш ерекшелік бар болсын (яғни, $n = 3$), бұлар j -шы О.Т.Б үшін x_{1j}, x_{2j}, x_{3j} мәндері бойынша анықталады, ал k -шы О.Т.Б үшін x_{1k}, x_{2k} және x_{3k} мәндерімен анықталсын. Яғни аталған екі таксономикалық бірліктерді үш өлшемді кеңістікте ара қашықтығы Δ_{jk} келесі өрнекпен анықталатын

$$\Delta_{jk}^2 = (x_{1j} - x_{1k})^2 + (x_{2j} - x_{2k})^2 + (x_{3j} - x_{3k})^2$$

екі нүкте (x_{1j}, x_{2j}, x_{3j}) және (x_{1k}, x_{2k}, x_{3k}) арқылы бейнелеуге болады. n ерекшелік жағдайында j және k таксономикалық бірліктер арасындағы қашықтыққа арналған жалпыланған өрнек:

$$\Delta_{jk}^2 = \sum_{i=1}^n (x_{ij} - x_{ik})^2.$$

x_{ij} және x_{ik} мәндері барлық i үшін бірдей жағдайда нүктелер арасындағы Δ_{jk} қашықтық нөлге тең болады. Кез келген айырмашылық осы мәнің өсуіне әкеледі, сондықтан қашықтықты айырмашылықтың индикаторы ретінде қарастырған жөн.

Тағы бір әдіс - корреляциялық коэффициенттерді қарастырғанда сипатталған матрицаның жолдарындағы деректерді нормалау. Егер z_{ij} нормаланған өлшемдеріу қолдансақ, қашықтыққа арналған формула келесі формуламен алмастырылады:

$$\Delta_{jk}^2 = \sum_{i=1}^n (z_{ij} - z_{ik})^2.$$

Δ_{jk} мәні тек z_{ij} - нақты өлшеу нәтижелеріне ғана байланысты емес, сондай-ақ салыстырмалы ерекшеліктерінің санына да (n) байланысты. Тек бір жағдайда, таксономикалық бірліктердің барлық жұптары арасындағы салыстыру n мәнінің бірдей мәніне негізделгенде бұл маңызды емес. Салыстыруды жеңілдету үшін, n -ге арнап әр түрлі мәндер пайдаланылған жағдайда, қашықтықтың орташа квадратын есептеген ыңғайлы:

$$d_{jk}^2 = \frac{\Delta_{jk}^2}{n}.$$

Сандық таксономия тұрғысынан таксономикалық қашықтықтар қолдануға қарапайым және ыңғайлы.

Сандық таксономиядағы талдау әдістері

Бұдан әрі, "қашық көршілер" әдісі немесе "толық байланыс" әдісі, "жақын көршілер" әдісі немесе "жалғыз байланыс" әдісі, "жұптасқан орташалар" әдісі сияқты иерархиялық талдаудың бірнеше қарапайым әдістері сипатталған. Әдістердің бірі - "қашық

көршілер" әдісі немесе "толық байланыс" әдісі. Нысандардың берілген жиынтығын "жақын көршілер" әдісін қолданып классификациялау қажет. Екі класс арасындағы қашықтық олардың жақын өкілдерінің арасындағы қашықтық ретінде анықталады. Ол үшін объектілер арасындағы қашықтық матрицасы есептеледі. Әрбір қадамда қашықтық матрицасында екі ең жақын объектілер арасындағы қашықтыққа сәйкес келетін ең аз мән анықталады. Нысандар біріктіріліп, жаңа нысан құрылады. Бұл процедура қарастырылатын барлық нысандар біріктірілмейінше қайталанады.

Мысалы, келесі қашықтық матрицасы берілген:

	1	2	3	4
1	0	2.06	4.03	6.32
2	2.06	0	4.12	2.25
3	4.03	4.12	0	3.50
4	6.32	2.25	3.50	0

1-қадам. Жіктеу критерийіне сәйкес, ең аз қашықтағы объектілер бірігеді: 1 және 2, 3 және 4. Бірігу қашықтығы 2,06 болды. Жаңа объектіні ескере отырып, қашықтық матрицасын қайта есептеу қажет:

	1,2	3	4
1,2	0	4.12	6.32
3	4.12	0	3.50
4	6.32	3.50	0

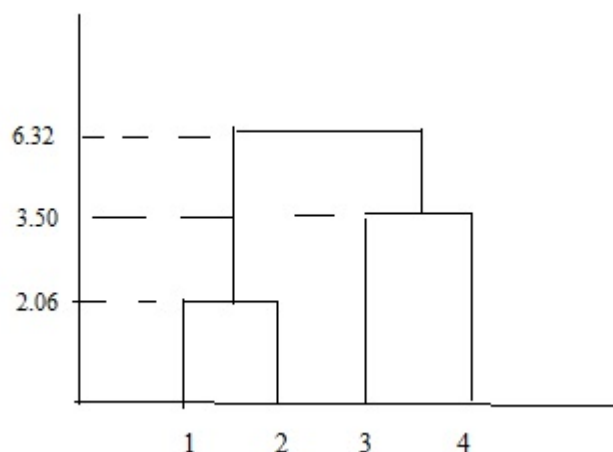
2-қадам. Осы қадамдағы объектілер: (1,2), 3, 4. Жаңа қашықтық матрицасына сәйкес 3 және 4 ең жақын объектілер. Біріктіру қашықтығы - 3,5. Жаңа объектіні ескере отырып, қашықтық матрицасын қайта есептеу қажет:

	1,2	3,4
1,2	0	6.32
3,4	6.32	0

3-қадам. Бұл қадамдағы объектілер: (1,2) және (3, 4). Олардың арасындағы қашықтық - 6.32, бұл 1-ші және 4-ші нысандар арасындағы қашықтық. Иерархияны қалыптастыру аяқталды. Ең жақын көршілер әдісімен жіктелу нәтижесі дендрограмма түрінде берілген:

"Жақын көршілер" әдісін қолданған кезде, объектілер арасындағы қашықтықты өлшеуде объектілер арасындағы қашықтыққа назар аудару керек. Осының негізінде қашықтықтардың бастапқы матрицасы қалыптасады, ол барлық жіктелу процесін анықтайды.

"Жақын көршілер" немесе "жалғыз байланыс" әдісі



1.1.1-сурет Жіктелу нәтижесі – дендрограмма

Иерархиялық талдаудың бұл әдісі объектілер арасындағы қашықтықты анықтағанда, қашықтық олардың ең жақын өкілдерінің арасындағы қашықтық ретінде алынуымен ерекшеленеді. Объектілердің берілген жиынтығын алыс көршілер әдісімен жіктеу қажет болсын. Екі объекті арасындағы қашықтық олардың жақын өкілдерінің арасындағы қашықтық ретінде анықталады.

Нысандар арасындағы ара қашықтық матрицасы есептеледі. Әрбір қадамда екі ең алыс объектілер арасындағы қашықтыққа сәйкес келетін қашықтық матрицасы анықталады. Объектілер біріктіріліп, жаңадан құрылады. Бұл процедура барлық нысандар біріктірілмейінше қайталанады. Мысалы, қашықтықтың келесі матрицасы берілген:

	1	2	3	4
1	0	2.06	4.03	6.32
2	2.06	0	2.5	4.12
3	4.03	2.5	0	2.24
4	6.32	4.12	2.24	0

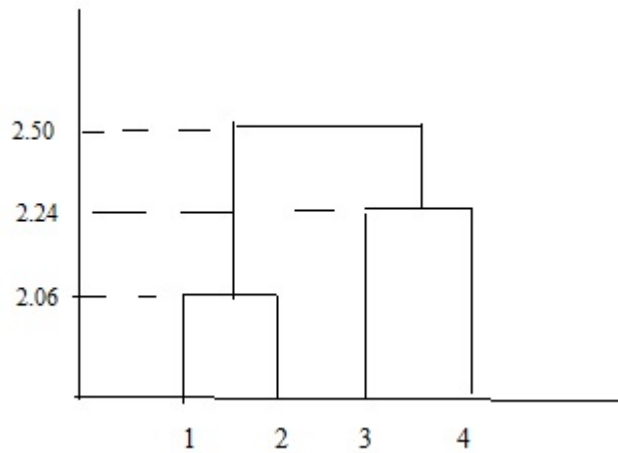
1-қадам. Жіктелу критерийіне сәйкес, объектілер арасындағы өзара байланыс, ең жақын өкілдерінің арасындағы қашықтық: 1 және 2, 3 және 4. Бірігудің қашықтығы 2,06 болды. Бұл қашықтық матрицасын есептеуде ескеріледі:

	1,2	3	4
1,2	0	2.5	4.12
3	2.5	0	2.24
4	4.12	2.24	0

2-қадам. Осы қадамдағы объектілер (1,2), 3, 4. Жаңа қашықтық матрицасына сәйкес 3 және 4 ең жақын. Бірігу қашықтығы - 2.24. Жаңа объектілерді ескере отырып, қашықтық матрицасын қайта есептеу қажет:

	1,2	3,4
1,2	0	2.5
3,4	2.5	0

3-қадам. Бұл қадамдағы нысандар: (1,2) және (3, 4). Олардың арасындағы қашықтық - 2.50. Бұл - 2-ші және 3-ші нысандардың арасындағы қашықтық. Иерархияны қалыптастыру аяқталды. "Жақын көрші" әдісімен жіктелу нәтижесі келесі дендрограмма түрінде беріледі.



1.1.2-сурет Жіктелу нәтижесі – дендрограмма

Жұптасқан орташалар әдісі

Бұл әдіс объектілер арасындағы қашықтық олардың жақын өкілдерінің аралықтары ретінде арифметикалық орташа мән алынатындығымен ерекшеленеді. K объектісі u және v объектілері біріктіру арқылы қалыптассын. Бұл кластерлер арасындағы қашықтық келесі формулаға сәйкес анықталады.

$$D((u, v), w) = \frac{D_{uw} + D_{vw}}{2}$$

Сонымен, иерархиялық классификациялау әдістері тек ұқсастық пен айырмашылықтарды бағалау әдістерімен ғана емес, сонымен қатар классификациялау алгоритмдерімен де ерекшеленетінін көруге болады.

Жаттығулар:

1. Статистикалық мәліметтер арқылы ұқсасатық матрицасын құрыңыз.
2. Келесі ерешеліктер матрицасы үшін "алыс көршілер" әдісін қолданып таксономикалық группалар құрыңыз, дендрограммасын келтіріңіз.

1	0	5.06	20.3	7.52	5.03
2	3.06	0	9.50	8.24	3.45
3	12.3	7.50	0	4.25	3.65
4	7.52	8.24	2.25	0	4.81
5	3.03	3.45	4.65	4.81	0

3. Келесі ерешеліктер матрицасы үшін "жақын көршілер" әдісін қолданып таксономикалық группалар құрыңыз, дендрограммасын келтіріңіз.

1	0	4.06	20.3	3.45	5.03
2	3.06	0	6.50	8.24	3.45
3	20.3	8.50	0	2.25	3.65
4	5.52	8.24	2.25	0	4.81
5	5.03	2.45	3.65	4.81	0

4. Келесі ерешеліктер матрицасы үшін "орташаланған байланыстар" әдісін қолданып таксономикалық группалар құрыңыз, дендрограммасын келтіріңіз.

1	0	9.06	15.3	5.52	5.03
2	3.06	0	7.50	8.24	3.45
3	20.3	9.50	0	2.25	3.65
4	7.52	7.24	2.25	0	4.81
5	5.03	3.45	5.65	4.81	0

5. Келесі ерешеліктер матрицасы үшін "жақын көршілер" әдісін қолданып таксономикалық группалар құрыңыз, дендрограммасын келтіріңіз.

1	0	2.06	40.3	6.32	2.08
2	2.06	0	3.50	4.12	5.43
3	4.03	3.50	0	2.25	3.65
4	6.32	4.12	2.25	0	4.81
5	2.08	5.43	3.65	4.81	0

6. Келесі ерешеліктер матрицасы үшін "орташаланған байланыстар" әдісін қолданып таксономикалық группалар құрыңыз, дендрограммасын келтіріңіз.

Әдебиеттер:

- (а) Бейли Н. Математика в биологии и медицине. Москва, Изд-во «Мир», 1970

- (b) С.А.Ляшко «Элементы теории динамических систем» Учебное пособие для студентов специальности «Математика», «Прикладная математика», «Прикладная информатика» Балашов, Изд-во «Николаев», 2005, 104С.
- (c) В.И. Арнольд «Обыкновенные дифференциальные уравнения» Ижевск: Удм.ГУ, 2000. – 368 с.