ДЕРЕКТЕРДІ ИНТЕЛЛЕКТУАЛДЫ ТАЛДАУ

Дәріс 11

Ықтималдық және сипаттамалық статистика

Оспан Ә.Ғ., PhD, доцент-зерттеуші

ҚазҰУ, Жасанды интеллект және Big Data кафедрасы

2026-2027 оқу жылы

Дәрістің мақсаты мен мазмұны

Мақсат және оқу нәтижелері

- Модель ықтималдығын бақыланған жиіліктен ажырату
- Математикалық күтімді, дисперсияны және стандартты ауытқуды есептеу
- Бақылаулар шашырауы мен таңдамалық орташалардың шашырауын ажырату

Дәріс мазмұны

1. **Кездейсоқ тәжірибе**
2. **Үлестірімдер**
3. **Математикалық күтім мен дисперсия**
4. **Қалыпты модель**
5. **Статистика және ОШТ**

Модель ықтималдығы және тәжірибедегі жиілік

Тәжірибеге дейін

- Ықтималдық тәжірибенің таңдалған моделін сипаттайды.
- Әділ сүйек үшін P(жұп) = 3/6 = 0.5.
- Нәтижелердің тең ықтималдығы негізделуі керек.

Тәжірибелер сериясынан кейін

- Жиілік нақты алынған нәтижелерге тәуелді.
- 20 нәтиженің 12-сі жұп болса, жиілік 0.6 болады.
- Шекті серия модельді дәл қайталауға міндетті емес.

Алдымен тәжірибені, оқиғаны және іріктеу механизмін анықтаңыз.

Таңдау тәсілі бақылаулардың салмағын өзгертеді

Тәжірибе	P(X=10)	P(X=40)	E(X)
Үш курстың бірін таңдау	2/3	1/3	20
60 студенттің бірін таңдау	1/3	2/3	30
Сұрақтағы айырмашылық	Курстар бойынша салмақ	Адамдар бойынша салмақ	Талдау бірліктері әртүрлі

Топ мүшесін таңдағанда үлкен топ жиірек кездеседі.

Кездейсоқ шама және оның үлестірімі

Дискретті модель

- Мәндер жиыны ақырлы немесе саналымды.
- p_i ықтималдықтары теріс емес; қосындысы 1-ге тең.
- Мысал: сатып алынған кітаптар саны.

Үздіксіз модель

- Ықтималдықтар мәндер аралықтарына қатысты.
- Тығыздық ықтималдықты аудан арқылы анықтайды.
- Мысал: көлік күту уақыты.

Санатты санмен жазу өздігінен сандық шкала бермейді.

Математикалық күтім — салмақталған орталық

$$E(X) = \sum(x_i * p_i)$$

Нені қосамыз

- Әрбір рұқсат етілген x_i мәнін p_i салмағымен.
- Жиі нәтиже салмақталған орталыққа көбірек әсер етеді.

Қалай түсіндіреміз

- Берілген модель қайталауларының орташа нәтижесі.
- Күтім міндетті түрде мүмкін нәтиже болуы шарт емес.

X: 1, 2, 3; p: 0.3, 0.4, 0.3. E(X)=2.

Үлестірімді тексеріп, оның орталығын есептейміз

Python мысалы

```
import numpy as np

x = np.array([1., 2., 3.])
p = np.array([0.3, 0.4, 0.3])
assert x.shape == p.shape
assert np.all(p >= 0) and np.isclose(p.sum(), 1)
mu = np.sum(x * p)
var = np.sum((x - mu) ** 2 * p)
print(mu, round(var, 3))
```

Нәтиже: 2.0 0.6

Шарттарды тексеру

- Массивтердің пішіні бірдей.
- Ықтималдықтар теріс емес және қосындысы бірге тең.

Детерминдік есептеу

- mu — берілген үлестірімнің орталығы.
- var — ауытқулар квадраттарының салмақталған мәні.

Дисперсия және стандартты ауытқу

$$Var(X) = E((X - E(X))^2); SD(X) = \sqrt{Var(X)}$$

Дисперсия

- Ауытқулар квадратталып, орташаланады.
- Бастапқы бірліктің квадратымен өлшенеді.
- Әрқашан теріс емес; тұрақты үшін нөлге тең.

Стандартты ауытқу

- Дисперсияның квадрат түбірі.
- Өлшем бірлігі X шамасымен бірдей.
- Орташа абсолюттік ауытқуға тең емес.

Кітап саны үшін: Var(X)=0.6; SD(X) шамамен 0.7746.

Түрлендірулер: орталық пен шашырау әртүрлі өзгереді

Түрлендіру	Математикалық күтім	Дисперсия
Y=aX+b	aE(X)+b	a**2 Var(X)
X+Y: жалпы жағдай	E(X)+E(Y)	Var(X)+Var(Y)+2Cov(X,Y)
X+Y: тәуелсіз шамалар	E(X)+E(Y)	Var(X)+Var(Y)
c тұрақтысы	c	0

Күтімнің сызықтылығы мен дисперсияларды қосудың шарттары әртүрлі.

Тығыздық және жинақталған ықтималдық

f(x) тығыздығы

- Қисық биіктігі нүктенің ықтималдығы емес.
- Аралық ықтималдығы қисық астындағы ауданға тең.
- Жалпы аудан 1-ге тең; f(x) 1-ден артық болуы мүмкін.

F(x) үлестірім функциясы

- F(x)=P(X<=x) сол жақтағы ықтималдықты жинақтайды.
- P(a<X<=b)=F(b)-F(a).
- ppf шекараны, sf оң жақ құйрықты табады.

Тығыздық гистограммасында биіктіктер қосындысын емес, аудандарды тексереді.

Қалыпты модель: орталық пен масштаб



Қалыпты тығыздықты оқу

- mu симметрия орталығын, sigma>0 масштабты анықтайды.
- z=(x-mu)/sigma ауытқуды SD бірлігімен өлшейді.
- Стандарттау кез келген пішінді қалыптыға айналдырмайды.
- Үш сигма ережесі қалыпты модельді талап етеді.

Қалыптылық — тексеруді қажет ететін модель жорамалы.

SciPy арқылы қалыпты құйрықтар мен квантильдер

Python мысалы

```
from scipy.stats import norm

upper = norm.sf(190, loc=182, scale=4)
middle = norm.cdf(186, loc=182, scale=4)
middle -= norm.cdf(178, loc=182, scale=4)
boundary = norm.ppf(0.975, loc=182, scale=4)
print(round(upper, 4), round(middle, 4))
print(round(boundary, 2))
```

Нәтиже: 0.0228 0.6827; квантиль 189.84

Тура есеп

- sf: 190-нан жоғары мәннің ықтималдығы.
- cdf айырмасы: 178-186 аралығының ықтималдығы.

Кері есеп

- ppf: жинақталған 0.975 үлесіне сәйкес шекара.
- loc — орташа мән; scale — стандартты ауытқу.

Деректер жиынын сипаттау және үлестірімді бағалау

Бақыланған ақырлы жиын

- Орташа мән, медиана, Q1, Q3 және IQR=Q3-Q1.
- Сипаттамалық дисперсия үшін ddof=0 қолдануға болады.
- Сан нақты қолдағы жолдарды сипаттайды.

Таңдама бойынша параметрді бағалау

- Әдетте s**2 квадраттар қосындысын n-1-ге бөледі.
- NumPy мен pandas-та ddof әдепкі мәндері әртүрлі.
- Іріктеу мен тәуелсіздікті бөлек негіздеу қажет.

football.csv ішіндегі Penalties — гол саны емес, дағды рейтингі.

ОШТ: бастапқы деректердің емес, орташа мәннің үлестірімі

E(X_bar)=mu; SE(X_bar)=sigma/sqrt(n)

Қайталама таңдамаларда не өзгереді

- Әр таңдама өзінің X_bar орташа мәнін береді.
- Стандартты қате осы орташа мәндердің шашырауын сипаттайды.

Шарттар мен шектеулер

- Тәуелсіздік, бірдей механизм, ақырлы дисперсия.
- «n>30 жеткілікті» деген әмбебап ереже жоқ.
- ОШТ бастапқы мәндерді қалыпты етпейді.

n көлемі 4 есе артса, стандартты қате 2 есе азаяды.

Сандардан қорытындыға көшудегі қателер

Модель

- Кез келген нәтижелер тізімін тең ықтималды деп санамау.
- Тәуелді дағдылардың дисперсияларын жай қоспау.

Сипаттама

- Тығыздық пен нүкте ықтималдығын шатастырмау.
- Бақылаулардың SD шамасы мен орташаның SE шамасын шатастырмау.

Деректер шекарасы

- Файл сипаттамасын барлық спортшыға жалпыламау.
- Үлкен n-ді тәуелсіздік дәлелі деп санамау.

Дәрістің негізгі нәтижелері

1. Ықтималдық модельді береді; жиілік шекті серияны сипаттайды.
2. Нысанды таңдау тәсілі салмақтар мен үлестірімді анықтайды.
3. Күтім — салмақталған орталық; дисперсия — ауытқудың орташа квадраты.
4. Үздіксіз ықтималдықтар аудандармен және cdf арқылы анықталады.
5. Орташаның стандартты қатесі 1/sqrt(n) заңымен азаяды.

Түсінуді тексеру және өзіндік жұмыс

1. E(X) неліктен мүмкін нәтиже болмауы мүмкін?
2. Клубты таңдау мен кездейсоқ ойыншының клубын таңдаудың айырмасы қандай?
3. Ковариациясыз дисперсияларды қашан қосуға болады?
4. Стандарттау неліктен қалыптылыққа кепілдік бермейді?
5. Бақылаулар SD шамасы мен орташаның SE шамасы несімен ерекшеленеді?

Өзіндік жұмыс

X=0,2,5 және ықтималдықтары 0.2,0.5,0.3 үшін E(X), Var(X) және Y=3X-4 сипаттамаларын табыңыз. Тікелей есептеумен тексеріңіз.

Құжаттама және бастапқы материалдар

Құжаттама

1. NumPy: кездейсоқ сандар генераторлары

https://numpy.org/doc/stable/reference/random/generator.html

2. SciPy: қалыпты үлестірім

https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.norm.html

3. pandas: сипаттамалық статистика

https://pandas.pydata.org/docs/user_guide/basics.html#descriptive-statistics

Дәрістің негізі

- Берілген 11-дәрістің ықтималдық және сипаттамалық статистика жөніндегі презентациясы
- Берілген 11-зертханалық жұмыстың материалы
- Берілген 11-зертханалық жұмыстың жауаптары бар материалы