МІНІСТЕРСТВО ВНУТРІШНІХ СПРАВ УКРАЇНИ
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ВНУТРІШНІХ СПРАВ
КРЕМЕНЧУЦЬКИЙ ЛЬОТНИЙ КОЛЕДЖ

Циклова комісія природничих дисциплін

ТЕКСТ ЛЕКЦІЇ

з навчальної дисципліни «Теорія ймовірностей і математична статистика»
обов'язкових компонент освітньо-професійної програми
першого (бакалаврського) рівня вищої освіти
Облік і аудит

за темою №5– Двомірні випадкові величини. Закон великих чисел. Центральна гранична теорема

Харків 2022

ЗАТВЕРДЖЕНО
Науково-методичною радою
Харківського національного
університету внутрішніх справ
Протокол від 30.08.2022 №8

СХВАЛЕНО
Методичною радою Кременчуцького
льотного коледжу Харківського
національного університету
внутрішніх справ
Протокол від 22.08.2022 №1

ПОГОДЖЕНО
Секцією Науково-методичної ради
ХНУВС з технічних дисциплін
Протокол від 29.08.2022 №8

Розглянуто на засіданні циклової комісії природничих дисциплін, протокол
від 10.08.2022 № 1

Розробник: доцент циклової комісії природничих дисциплін, к.ф. –м.н.
Семенов В.О.

Рецензенти:
1. Доцент кафедри автомобілі та трактори Кременчуцького національного університету імені Михайла Остроградського, кандидат технічних наук
Черниш А.А.
2. Ст.викладач циклової комісії економіки та управління Кременчуцького льотного коледжу Харківського національного університету внутрішніх справ, кандидат економічних наук Цимбалістова О.А.

План лекції

1. Закони розподілу та числові характеристики дискретних двовимірних випадкових величин.
2. Закон великих чисел.
3. Центральна гранична теорема.

Рекомендована література:

Основна

1. Барковський В. В. Теорія ймовірностей та математична статистика ; 5-те видання : навчальний посібник / В. В. Барковський, Н. В. Барковська, О. К. Лопатін. – К. : Центр учбової літератури, 2010. – 424 с.
2. Волощенко А. Б. Теорія ймовірностей і математична статистика : навчаль-но-методичний посібник для самостійного вивчення дисципліни / А. Б. Воло-щенко, І. А. Джалладова. – К. : КНЕУ, 2003. – 256 с.
3. Жильцов О.Б. Теорія ймовірностей та математична статистика у прикладах і задачах: навч. посіб. для студ. вищ. навч. закл. / О.Б. Жильцов ; за ред. Г.О. Михаліна. — К. : Київ. ун-т ім. Б. Грінченка, 2015. — 336 с.
4. Конспект лекцій.

Інформаційні ресурси в інтернеті

5. Жильцов О.Б. Теорія ймовірностей та математична статистика у прикладах і задачах: навч. посіб. для студ. вищ. навч. закл. / О.Б. Жильцов ; за ред. Г.О. Михаліна. — К. : Київ. ун-т ім. Б. Грінченка, 2015. — 336 с. [https://yandex.ua/search/?lr=143&oprnd=9346860898&text=жильцов%20теорія%20ймовірностей

Текст лекції

1. Закони розподілу та числові характеристики дискретних двовимірних випадкових величин.

Якщо задано дві випадкові величини X і Y, то вважають, що задано *систему двох випадкових величин* (X, Y), або *двовимірну випадкову величину.*При вивченні системи випадкових величин необхідно враховувати залежність між ними. *Функцією розподілу* двовимірної випадкової величини (X, Y) називають функцію двох змінних

$$F(x, y) = P(X < x, Y < y),$$

де $X < x$, $Y < y$ означає подію попадання випадкової точки (X, Y) у нескінченний квадрант з вершиною в точці (x, y) (лівіше й нижче від точки (x, y)).

Функція розподілу двовимірної випадкової величини має такі властивості.

1. $\lim\limits_{\substack{x\to+\infty\\y\to+\infty}} F(x,y)=1;\quad \lim\limits_{x\to x_0-0} F(x,y)=F(x_0,y);\quad \lim\limits_{y\to y_0-0} F(x,y)=F(x,y_0);$

$\lim\limits_{\substack{x\to x_0-0\\y\to y_0-0}} F(x,y)=F(x_0,y_0);\quad \lim\limits_{x\to-\infty} F(x,y)=\lim\limits_{y\to-\infty} F(x,y)=\lim\limits_{\substack{x\to-\infty\\y\to-\infty}} F(x,y)=0.$

2. $F(x, y)$ неспадна по кожній змінній, тобто:

а) якщо $y_1 < y_2$, то $F(x, y_1) \le F(x, y_2)$ для всіх $x \in R$;

б) якщо $x1 < x2$, то $F(x1, y) \le F(x2, y)$ для всіх $y \in R$.

3. Якщо випадкова величина X має функцію розподілу $F_X(y)$, а випадкова величина Y — функцію розподілу $F_Y(x)$, то

$$\lim_{x \to +\infty} F(x, y) = F_X(y); \quad \lim_{y \to +\infty} F(x, y) = F_Y(x).$$

4. Ймовірність влучення випадкової точки (X, Y) у прямокутник [a; b)×[c; d) зі сторонами, які паралельні координатним осям, обчислюється за формулою

$$P(a \le X < b;\ c \le Y < d) = F(b, d) - F(a, d) - F(b, c) + F(a, c).$$

Як і одновимірні, двовимірні випадкові величини можна розподілити на два класи: *дискретні* та *неперервні*. Двовимірна випадкова величина (X, Y) називається *дискретною*, якщо і випадкова величина X, і випадкова величина Y є дискретними випадковими величинами. При цьому існують значення x_j, $j = 1, 2, \dots$, та значення y_i, $i = 1, 2, \dots$, для яких

$$P(X = x_j\ ;\ Y = y_i) = Pi,j \ge 0 \text{ та } \sum_i \sum_j P_{ij} = 1$$

У випадку, коли $1 \le j \le n$, $1 \le i \le m$, дискретну двовимірну випадкову величину (X, Y) зручно задавати за допомогою таблиці.

X / Y	x_1	x_2	…	x_n
y_1	p_{11}	p_{12}	…	p_{1n}
y_2	p_{21}	p_{22}	…	p_{2n}
…	…	…	…	…
y_m	p_{m1}	p_{m2}	…	p_{mn}

Для дискретної двовимірної випадкової величини (X, Y) функція розподілу набуває вигляду суми

$$F(x, y) = \sum_{(i,j):\ \begin{cases} x_j < x; \\ y_i < y. \end{cases}} p_{ij},$$

тобто обчислення сум виконується за всіма наборами (i, j), для яких одночасно $x_j < x$ і $y_i < y$. Користуючись наведеною вище таблицею, легко визначити розподіли кожної випадкової величини окремо за допомогою формул:

$$P(X = x_j) = p_{1j} + p_{2j} + \dots + p_{mj}, \quad j = 1,2, \dots;$$
$$P(Y = y_i) = p_{i1} + p_{i2} + \dots + p_{im}, \quad i = 1,2, \dots.$$

Якщо випадкові величини X і Y залежні, то для знаходження розподілу двовимірної випадкової величини (X, Y) недостатньо знати закони розподілу кожної з випадкових величин X і Y. *Умовним законом розподілу* називається

розподіл однієї випадкової величини, знайдений за умови, що інша випадкова величина системи набула деякого фіксованого значення.

Числові характеристики умовних законів розподілу обчислюють як і раніше.

Для опису рівня залежності випадкових величин X і Y використовують такі числові характеристики як *коваріація* та *коефіцієнт кореляції*.

Коваріацією (*кореляційним моментом*) системи випадкових величин (X, Y) називають число

$$\operatorname{cov}(X,Y) = \sum_i \sum_j \big(x_i - M(x_i)\big)\big(y_i - M(y_i)\big)\, p_{ij}.$$

Коефіцієнтом кореляції системи випадкових величин X і Y називають число

$$r(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sigma(X)\sigma(Y)}.$$

Якщо X і Y — незалежні випадкові величини, то $\operatorname{cov}(X, Y) = r(X, Y) = 0$.

Обернене твердження неправильне. Існують залежні випадкові величини, для яких коваріація та коефіцієнт кореляції дорівнюють нулю. Для довільних випадкових величин

$$-1 \le r(X, Y) \le 1.$$

Чим ближче $|r(X, Y)|$ до одиниці, тим сильнішим вважають рівень залежності між X і Y. Якщо $r(X, Y) = 0$, то випадкові величини називаються *некорельованими*.

***Приклад* 1.** Із коробки, у якій міститься 3 червоні й 2 сині кульки, навмання без повторень виймають послідовно кульки до першої появи синьої кульки. Далі кульки виймають до першої появи червоної кульки або до закінчення кульок у коробці. Необ-хідно описати закон розподілу ймовірностей системи випадкових величин (X, Y), де X —кількість червоних кульок, узятих із коробки до першої появи синьої кульки; Y — кількість синіх кульок, узятих із коробки до першої появи червоної кульки або до закінчення кульок після того, як перша синя кулька булла вийнята з коробки. Скласти окремі закони розподілу для випадкових величин X і Y.

Розв'язання. Випадкова величина X може набувати значень 0, 1, 2, 3, а випадкова величина Y — значень 0, 1. Обчислимо p_{ij}, $i = 0,1,2,3$, $j = 0,1$ — ймовірності того, що $X = i$, $Y = j$:

$$p_{0,0} = (x = 0, y = 0) = \frac{2}{5} \cdot \frac{3}{4} \cdot \frac{3}{3} = \frac{6}{20};$$

$$p_{0,1} = (x = 0, y = 1) = \frac{2}{5} \cdot \frac{3}{4} = \frac{6}{20};$$

$$p_{1,0} = (x = 1, y = 0) = \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{2}{3} = \frac{4}{20};$$

$$p_{1,1} = (x = 1, y = 1) = \frac{2}{20};$$

$$p_{2,0} = (x = 2, y = 0) = \frac{2}{20};$$

$$p_{2,1} = (x = 2, y = 1) = \frac{2}{20};$$

$$p_{3,0} = (x = 3, y = 0) = 0;$$

$$p_{3,1} = (x = 3, y = 1) = \frac{2}{20}.$$

Отримані значення запишемо в таблицю розподілу ймовірностей системи випадкових величин (*X*, *Y*):

X / *Y*	0	1	2	3	$\sum$
0	$\frac{6}{20}$	$\frac{4}{20}$	$\frac{2}{20}$	0	$\frac{12}{20}$
1	$\frac{2}{20}$	$\frac{2}{20}$	$\frac{2}{20}$	$\frac{2}{20}$	$\frac{8}{20}$
$\sum$	$\frac{8}{20}$	$\frac{6}{20}$	$\frac{4}{20}$	$\frac{2}{20}$	1

Із отриманої таблиці досить легко записати безумовні закони розподілу для випадкових величин *X* і *Y*:

x	0	1	2	3
p	$\frac{8}{20}$	$\frac{6}{20}$	$\frac{4}{20}$	$\frac{2}{20}$

y	0	1
p	$\frac{12}{20}$	$\frac{8}{20}$

2. Закон великих чисел.

Граничні теореми теорії ймовірностей можна умовно розбити на два класи*: закони великих чисел* і *центральні граничні теореми*. Закони великих чисел математично описують стійкість середніх значень масових випадкових

явищ. Історично першим законом великих чисел була така теорема:

Теорема Я. Бернуллі. Нехай $k = k_n$ — кількість появ події A в серії з n послідовних незалежних випробувань, у кожному з яких подія A відбувається з ймовірністю p. Тоді для довільного $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) = 1.$$

Іншими словами, з ймовірністю, як завгодно близькою до одиниці, можна стверджувати, що при досить великій кількості n незалежних послідовних випробувань відносна частота $\frac{k}{n}$ появи події A скоріше за все досить мало відрізняється від ймовірності появи події A.

Ця теорема, зокрема, пояснює, що, коли при багаторазовому підкиданні мо-нети кількість гербів становить приблизно половину від загальної кількості підкидань, тоді доцільно вважати ймовірність випадання герба (та цифр) рівною 0,5. Узагальненням теореми Бернуллі є

Теорема П. Л. Чебишева. Нехай Xi — послідовність попарно незалежних випадкових величин зі скінченними математичними сподіваннями $M(Xi)$ і обмеженими дисперсіями $D(Xi) \le c_0$. Тоді для довільного $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} M(X_i)\right| < \varepsilon\right) = 1.$$

В обох теоремах суттєво використано нерівність Чебишева, яка, втім, має й самостійну цінність:

Лема (нерівність Чебишева). Нехай X — довільна випадкова величина з дисперсією $D(X) < \infty$ і математичним сподіванням $M(X)$. Тоді для довільного $\varepsilon > 0$ справджується нерівність:

$$P\left(|X - M(X)| < \varepsilon\right) \ge 1 - \frac{D(X)}{\varepsilon^2}.$$

Нерівність Чебишева можна записати також у формі

$$P\left(|X - M(X)| \le \varepsilon\right) \le \frac{D(X)}{\varepsilon^2}.$$

3. Центральна гранична теорема.

Крім законів великих чисел, описаних у теоремах Бернуллі та Чебишова, спостерігається ще одна цікава закономірність, яка полягає в тому, що при великій кількості випадкових доданків, кожний з яких вносить лише невеликий внесок у загальну суму, розподіл кожного з випадкових доданків не впливає на сумарний результат. Більш точне твердження сформульоване в теоремі, яку довів видатний російський математик О.М. Ляпунов.

Центральна гранична теорема. Нехай Xi — незалежні однаково розподілені випадкові величини з математичним сподіванням $M(Xi) = a$ і дисперсією $D(Xi) = \sigma^2$. Тоді при великому n розподіл суми $X_1 + X_2 + \ldots + Xn = X$ близький до нормального розподілу. При цьому

$$P(\alpha \le X \le \beta) \approx \Phi\left(\frac{\beta - na}{\sigma\sqrt{n}}\right) - \Phi\left(\frac{\alpha - na}{\sigma\sqrt{n}}\right).$$

З останніх рівностей випливає, що ймовірність p події можна оцінити, вико-ристовуючи експериментальні результати. Справді, якщо в кожному випробуванні подія A відбувається з ймовірністю p, і в серії з n незалежних вип.-робувань досліджувана подія A відбувається k разів, то статистична частота появи події A дорівнює $\frac{k}{n}$. Ця частота є випадковою величиною X, яка є сумою незалежних випадкових величин X_k, що набувають значень $\frac{1}{n}$ з ймо-вірністю p, коли подія A відбувається у k-му випробуванні, і 0 з ймовірністю q = 1 — p в іншому разі, k = 1,2, …, n. Для таких випадкових величин X_k математичне сподівання $M(X_k) = \frac{p}{n}$, а дисперсія $\sigma^2 = D(X_k) = \frac{p(1-p)}{n^2}$, $k = 0,1,2,...,n.$ Тоді при досить великих значеннях n ($n \ge 30$) виконується рівність

$$P\left(\left|\frac{k}{n} - p\right| < \delta\right) \approx 2\Phi\left(\frac{\delta}{\sigma\sqrt{n}}\right) = 2\Phi\left(\delta\sqrt{\frac{n}{p(1-p)}}\right).$$

Приклад 2. Використовуючи нерівність Чебишева, оцінити ймовірність поя-ви події A, яка полягає у тому, що випадкова величина X набуде значення, яке відрізнятиметься від математичного сподівання M (X) на величину, що не перевищує потроєного середнього квадратичного відхилення. Чи зміниться відповідь, якщо відомо, що випадкова величина X має нормальний розподіл?

Розв'язання. За нерівністю Чебишева

$$P(A) = P\left(\left|X - M(X)\right| < 3\sigma(X)\right) \ge 1 - \frac{D(X)}{(3\sigma(X))^2} = 1 - \frac{D(X)}{9D(X)} = \frac{8}{9}.$$

Отже, $P(A) \ge \frac{8}{9} \approx 0,89.$ Якщо X має нормальний розподіл, то

$$P(A) = P\left(\left|X - M(X)\right| < 3\sigma(X)\right) \approx 2\Phi\left(\frac{3\sigma(X)}{\sigma(X)}\right) = 2\Phi(3) \approx 0,9973.$$