

Залікова робота з курсу “Дискретна теорія ймовірностей”

Студента групи МП-31 Захарова Дмитра Олеговича

7 грудня 2023 р.

Варіант 3.

Завдання 1. Біномінальний закон розподілу. Приклад.

Означення

Нехай ми проводимо N незалежних спостережень, в кожному з яких може відбутися подія з деякою ймовірністю $\theta \in [0, 1]$ або не відбутися з ймовірністю $1 - \theta$.

Розглядаємо випадкову величину ξ яка дорівнює кількості випадіння події. Тоді, біномінальний розподіл можна записати як:

$$\text{Bin}(k \mid N, \theta) \equiv \mathbb{P}(\xi = k) \triangleq \binom{N}{k} \theta^k (1 - \theta)^{N-k}, \quad k \in \{0, \dots, N\}$$

Цю формулу можемо зрозуміти наступним чином: ймовірність отримати k успіхів θ^k , а $N - k$ неуспіхів $(1 - \theta)^{N-k}$. Проте, ми можемо отримати k успіхів з N спроб $\binom{N}{k}$ способами. Звідси і така формула розподілу.

Якщо випадкова величина ξ підпорядковується біноміальному розподілу, то це позначається як $\xi \sim \text{Bin}(N, \theta)$.

Також, є інший розподіл доволі пов'язаний з біноміальним. Його називають *мультиноміальним* законом розподілу. Нехай ми спостерігаємо N разів мульткласову подію, котра може приймати C різних значень з відповідними ймовірностями $\boldsymbol{\theta} = [\theta_1, \dots, \theta_C]^\top$ (звичайно, що при цьому маємо вимагати $\sum_{i=1}^C \theta_i = 1$). Тобто наш розподіл складається з подій $\zeta_n \sim \text{Cat}(\boldsymbol{\theta})$, де $\text{Cat}(y \mid \boldsymbol{\theta}) \triangleq \prod_{c=1}^C \theta_c^{\mathbf{1}(y=c)}$. Позначимо $\mathbf{y} := [N_1, \dots, N_C]^\top$ – вектор, $c^{\text{й}}$ елемент котрого позначає кількість випадіння $c^{\text{ої}}$ події, тобто $y_c = N_c := \sum_{n=1}^N \mathbf{1}(\zeta_n = c)$. Тоді, мультиноміальний розподіл має вигляд:

$$\mathcal{M}(\mathbf{y} \mid N, \boldsymbol{\theta}) \triangleq \binom{N}{y_1 \dots y_C} \prod_{c=1}^C \theta_c^{y_c} = \binom{N}{N_1 \dots N_C} \prod_{c=1}^C \theta_c^{N_c}$$

де ми позначили

$$\binom{N}{N_1 \dots N_C} \triangleq \frac{N!}{N_1! N_2! \dots N_C!}$$

Властивості біноміального розподілу

Нехай $\xi \sim \text{Bin}(N, \theta)$. Тоді:

Властивість 1. Математичне сподівання $\mathbb{E}[\xi] = N\theta$

Доведення. За означенням, $\xi \triangleq \sum_{n=1}^N \zeta_n$, де $\zeta_n \sim \text{Ber}(\theta)$ – незалежні випадкові величини, що підпорядковуються розподілу Бернуллі. Тоді, користуючись лінійністю математичного сподівання,

$$\mathbb{E}[\xi] = \mathbb{E}\left[\sum_{n=1}^N \zeta_n\right] = \sum_{n=1}^N \mathbb{E}[\zeta_n] = \sum_{n=1}^N \theta = N\theta \quad \blacksquare$$

Коментар. Можна було доводити за означенням математичного сподівання $\mathbb{E}[\xi] = \sum_{n=0}^N n\mathbb{P}(\xi = n)$, але доведення було б доволі громіздким.

Властивість 2. Дисперсія $\mathbb{V}[\xi] = n\theta(1 - \theta)$.

Доведення. Оскільки $\xi \triangleq \sum_{n=1}^N \zeta_n$, $\zeta_n \sim \text{Ber}(\theta)$ і $\{\zeta_n\}_{n=1}^N$ попарно незалежні, то

$$\mathbb{V}[\xi] = \mathbb{V}\left[\sum_{n=1}^N \zeta_n\right] = \sum_{n=1}^N \mathbb{V}[\zeta_n] = \sum_{n=1}^N \theta(1 - \theta) = N\theta(1 - \theta) \blacksquare$$

Наближення біноміального розподілу

Наведемо деякі корисні теореми, котрі можуть допомогти наближено оцінити значення ймовірностей для великої кількості спостережень. Оскільки це не безпосередньо тема питання, наведемо їх без доведень.

Теорема Пуассона. Нехай $\zeta_n \sim \text{Ber}(\theta_n)$, тобто на кожному спостереженні ймовірність θ_n , взагалі кажучи, може бути різною. Нехай при цьому $n\theta_n \xrightarrow{n \rightarrow \infty} \alpha \in \mathbb{R}$, тоді

$$\lim_{N \rightarrow \infty} \mathbb{P}(\xi = k) = \frac{\alpha^k}{k!} e^{-\alpha}, \quad k \in \mathbb{N} \cup \{0\}$$

Інтегральна теорема Муавра-Лапласа. Нехай $\xi_N \sim \text{Bin}(N, \theta)$. Тоді

$$\lim_{N \rightarrow \infty} \mathbb{P}\left(\alpha \leq \frac{\xi_N - N\theta}{\sqrt{N\theta(1 - \theta)}} \leq \beta\right) = \frac{1}{\sqrt{2\pi}} \int_{[\alpha, \beta]} e^{-\frac{x^2}{2}} dx$$

Зауваження до інтегральної теореми Муавра-Лапласа. Насправді, ця властивість випливає з того факту, що для достатньо великих $N \gg 1$, біноміальний розподіл приблизно дорівнює нормальному розподілу $\mathcal{N}(N\theta, N\theta(1 - \theta))$, густина котрого $p(x) = \frac{1}{\sqrt{2\pi N\theta(1 - \theta)}} \exp\left\{-\frac{(x - N\theta)^2}{2N\theta(1 - \theta)}\right\}$ (окрім стандартного доведення, що наводилось у теорії).

Приклад

Нехай Дмитро ходить на пару з дискретної теорії ймовірності з ймовірністю 0.7. Якщо в семестрі 13 пар з цього предмету, то запишіть:

1. Ймовірність того, що Дмитро відвідає рівно k пар.

2. Найбільш ймовірне значення кількості пар, які відвідає Дмитро.
3. Дисперсію кількості пар, які відвідає Дмитро.

Розглядаємо випадкову величину $\xi \sim \text{Bin}(13, 0.7)$, що позначає кількість пар, що відвідав Дмитро. Для першого пункту запишемо за означенням:

$$\mathbb{P}(\xi = k) = \text{Bin}(k \mid 13, 0.7) = \binom{13}{k} 0.7^k \cdot 0.3^{13-k}$$

Для другого знаходимо математичне сподівання: $\mathbb{E}[\xi] = 13 \cdot 0.7 = 9.1$, тобто скоріше за все, Дмитро відвідає близько 9 пар. Для дисперсії ж маємо $\mathbb{V}[\xi] = 13 \cdot 0.7 \cdot 0.3 = 2.73$.

Завдання 2.

Умова. Дано таблицю розподілу двовимірного випадкового вектору $[\xi, \eta]^\top$:

—	$\eta = -1$	$\eta = 0$	$\eta = 2$
$\xi = 0$	0.20	0.01	0.03
$\xi = 1$	0.01	0.03	0.03
$\xi = 3$	0.08	0.2	x

1. Знайти невідоме значення параметру x , знайти ймовірності $\mathbb{P}(\xi < \eta)$, $\mathbb{P}(\xi \leq \eta)$.
2. Знайти умовний закон розподілу η за умови, що $\xi = 3$ та умовний закон розподілу ξ за умови, що $\eta = 0$.
3. Побудувати функцію розподілу дискретного випадкового вектору $[\xi, \eta]^\top$.
4. Знайти коефіцієнт кореляції випадкових величин ξ та η . Чи будуть величини ξ та η незалежними? Якщо ні, то побудуйте таблицю розподілу двовимірного випадкового вектору $[\xi, \eta]^\top$ з незалежними компонентами.

Розв'язання.

Пункт 1. Позначимо через X та Y набір можливих значень ξ та η , відповідно. Тоді, щоб знайти x , скористаємося тим фактом, що

$$\sum_{(x,y) \in X \times Y} \mathbb{P}(\xi = x, \eta = y) = 1$$

Отже, $0.59 + x = 1.0 \implies \boxed{x = 0.41}$. Тепер знайдемо ймовірності з умови:

$$\begin{aligned} \mathbb{P}(\xi < \eta) &= \sum_{\substack{(x,y) \in X \times Y: \\ x < y}} \mathbb{P}(\xi = x, \eta = y) \\ &= \mathbb{P}(\xi = 0, \eta = 2) + \mathbb{P}(\xi = 1, \eta = 2) = 0.03 + 0.03 = \boxed{0.06} \end{aligned}$$

$$\mathbb{P}(\xi \leq \eta) = \underbrace{\mathbb{P}(\xi < \eta)}_{=0.06} + \mathbb{P}(\xi = \eta) = 0.06 + \mathbb{P}(\xi = 0, \eta = 0) = \boxed{0.07}$$

Пункт 2. Отже, треба знайти $\mathbb{P}(\eta \mid \xi = 3)$ та $\mathbb{P}(\xi \mid \eta = 0)$. Почнемо з першого:

$$\begin{aligned} \mathbb{P}(\eta = -1 \mid \xi = 3) &= \frac{\mathbb{P}(\xi = 3, \eta = -1)}{\mathbb{P}(\xi = 3)} = \frac{\mathbb{P}(\xi = 3, \eta = -1)}{\sum_{y \in Y} \mathbb{P}(\xi = 3, \eta = y)} = \frac{0.08}{0.69} = \frac{8}{69} \\ \mathbb{P}(\eta = 0 \mid \xi = 3) &= \frac{\mathbb{P}(\xi = 3, \eta = 0)}{\sum_{y \in Y} \mathbb{P}(\xi = 3, \eta = y)} = \frac{0.2}{0.69} = \frac{20}{69} \\ \mathbb{P}(\eta = 2 \mid \xi = 3) &= \frac{\mathbb{P}(\xi = 3, \eta = 2)}{\sum_{y \in Y} \mathbb{P}(\xi = 3, \eta = y)} = \frac{41}{69} \end{aligned}$$

Отже, наводимо розподіл знизу:

y	-1	0	2
$\mathbb{P}(\eta = y \mid \xi = 3)$	$\frac{8}{69}$	$\frac{20}{69}$	$\frac{41}{69}$

Аналогічно прописуємо для $\mathbb{P}(\xi \mid \eta = 0)$:

$$\begin{aligned}\mathbb{P}(\xi = 0 \mid \eta = 0) &= \frac{\mathbb{P}(\xi = 0, \eta = 0)}{\sum_{x \in X} \mathbb{P}(\xi = x, \eta = 0)} = \frac{0.01}{0.24} = \frac{1}{24} \\ \mathbb{P}(\xi = 1 \mid \eta = 0) &= \frac{\mathbb{P}(\xi = 1, \eta = 0)}{\sum_{x \in X} \mathbb{P}(\xi = x, \eta = 0)} = \frac{0.03}{0.24} = \frac{3}{24} = \frac{1}{8} \\ \mathbb{P}(\xi = 3 \mid \eta = 0) &= \frac{\mathbb{P}(\xi = 3, \eta = 0)}{\sum_{x \in X} \mathbb{P}(\xi = x, \eta = 0)} = \frac{0.2}{0.24} = \frac{20}{24} = \frac{5}{6}\end{aligned}$$

Тоді таблиця розподілу:

x	0	1	3
$\mathbb{P}(\xi = x \mid \eta = 0)$	$\frac{1}{24}$	$\frac{1}{8}$	$\frac{5}{6}$

Пункт 3. Будуємо функцію розподілу:

$$F_{(\xi, \eta)}(x, y) = \begin{cases} 0, & x \leq 0 \vee y \leq -1 \\ 0.20, & 0 < x \leq 1 \wedge -1 < y \leq 0 \\ 0.20 + 0.01, & 0 < x \leq 1 \wedge 0 < y \leq 2 \\ 0.20 + 0.01 + 0.03, & 0 < x \leq 1 \wedge y > 2 \\ 0.20 + 0.01, & 1 < x \leq 3 \wedge -1 < y \leq 0 \\ 0.20 + 0.01 + 0.01 + 0.03, & 1 < x \leq 3 \wedge 0 < y \leq 2 \\ 0.20 + 0.01 + 0.03 + 0.01 + 0.03 + 0.03, & 1 < x \leq 3 \wedge y > 2 \\ 0.20 + 0.01 + 0.08, & x > 3 \wedge -1 < y \leq 0 \\ 0.20 + 0.01 + 0.01 + 0.03 + 0.08 + 0.20, & x > 3 \wedge 0 < y \leq 2 \\ 1, & x > 3 \wedge y > 2 \end{cases}$$

Або, якщо порахуємо:

$$F_{(\xi, \eta)}(x, y) = \begin{cases} 0.00, & (x, y) \in (-\infty, 0] \times (-\infty, -1] \\ 0.20, & (x, y) \in (0, 1] \times (-1, 0] \\ 0.21, & (x, y) \in (0, 1] \times (0, 2] \\ 0.24, & (x, y) \in (0, 1] \times (2, +\infty) \\ 0.21, & (x, y) \in (1, 3] \times (-1, 0] \\ 0.25, & (x, y) \in (1, 3] \times (0, 2] \\ 0.31, & (x, y) \in (1, 3] \times (2, +\infty) \\ 0.29, & (x, y) \in (3, +\infty) \times (-1, 0] \\ 0.53, & (x, y) \in (3, +\infty) \times (0, 2] \\ 1.00, & (x, y) \in (3, +\infty) \times (2, +\infty) \end{cases}$$

Пункт 4. За означенням, коефіцієнт кореляції:

$$\text{corr}[\xi, \eta] \triangleq \frac{\text{Cov}[\xi, \eta]}{\sqrt{\mathbb{V}[\xi]\mathbb{V}[\eta]}}$$

Причому, для обрахунку коваріації використовуємо формулу

$$\text{Cov}[\xi, \eta] = \mathbb{E}[\xi\eta] - \mathbb{E}[\xi]\mathbb{E}[\eta]$$

Отже, знаходимо математичні сподівання та дисперсії ξ, η , а потім $\mathbb{E}[\xi\eta]$. Отже, за означенням,

$$\begin{aligned} \mathbb{E}[\xi] &\triangleq \sum_{x \in X} x \mathbb{P}(\xi = x) = \sum_{x \in X} x \sum_{y \in Y} \mathbb{P}(\xi = x, \eta = y) \\ &= 1 \cdot (0.01 + 0.03 + 0.03) + 3 \cdot (0.08 + 0.2 + 0.41) = 2.14 \\ \mathbb{E}[\eta] &\triangleq \sum_{y \in Y} y \mathbb{P}(\eta = y) = \sum_{y \in Y} y \sum_{x \in X} \mathbb{P}(\xi = x, \eta = y) \\ &= -1 \cdot (0.2 + 0.01 + 0.08) + 2 \cdot (0.03 + 0.03 + 0.41) = 0.65 \end{aligned}$$

Для знаходження дисперсії знайдемо $\mathbb{E}[\xi^2]$ та $\mathbb{E}[\eta^2]$:

$$\begin{aligned}\mathbb{E}[\xi^2] &= \sum_{x \in X} x^2 \mathbb{P}(\xi = x) = 1^2 \cdot 0.07 + 3^2 \cdot 0.69 = 6.28 \\ \mathbb{E}[\eta^2] &= \sum_{y \in Y} y^2 \mathbb{P}(\eta = y) = (-1)^2 \cdot 0.29 + 2^2 \cdot 0.47 = 2.17\end{aligned}$$

Нарешті, математичне сподівання :

$$\mathbb{E}[\xi\eta] = \sum_{(x,y) \in X \times Y} xy \mathbb{P}(\xi = x, \eta = y) = -0.01 - 0.08 \cdot 3 + 2 \cdot 0.03 + 6 \cdot 0.41 = 2.27$$

Отже, ми готові все поєднувати. Маємо:

$$\text{Cov}[\xi, \eta] = \mathbb{E}[\xi\eta] - \mathbb{E}[\xi]\mathbb{E}[\eta] = 2.27 - 2.14 \cdot 0.65 = 0.879$$

Тепер рахуємо дисперсії:

$$\begin{aligned}\mathbb{V}[\xi] &= \mathbb{E}[\xi^2] - \mathbb{E}[\xi]^2 = 6.28 - 2.14^2 = 1.7004 \\ \mathbb{V}[\eta] &= \mathbb{E}[\eta^2] - \mathbb{E}[\eta]^2 = 2.17 - 0.65^2 = 1.7475\end{aligned}$$

Отже коефіцієнт кореляції:

$$\text{corr}[\xi, \eta] \triangleq \frac{0.879}{\sqrt{1.7004 \cdot 1.7475}} \approx \frac{0.879}{1.724} \approx 0.51$$

Перевіримо, чи є ξ та η незалежними. Для цього має виконуватись наступне твердження:

$$\forall (x, y) \in X \times Y : \mathbb{P}(\xi = x, \eta = y) = \mathbb{P}(\xi = x)\mathbb{P}(\eta = y)$$

Візьмемо $(x, y) = (0, -1)$. В такому разі, $\mathbb{P}(\xi = 0, \eta = -1) = 0.20$ з таблиці. Проте,

$$\mathbb{P}(\xi = 0)\mathbb{P}(\eta = -1) = 0.24 \cdot 0.29 = 0.0696 \neq 0.20$$

Отже, $[\xi, \eta]^\top$ є залежними.

Тому побудуємо таблицю розподілу з незалежними компонентами. Для цього візьмемо розподіл $p(\xi = x)$ та $p(\eta = y)$ та складемо розподіл $[\xi, \eta]^\top$ таким чином, щоб $\mathbb{P}(\xi = x, \eta = y) = \mathbb{P}(\xi = x)\mathbb{P}(\eta = y)$.

Розподіл $\mathbb{P}(\xi = x)$:

x	0	1	3
$\mathbb{P}(\xi = x)$	0.24	0.07	0.69

Розподіл $\mathbb{P}(\eta = y)$:

y	-1	0	2
$\mathbb{P}(\eta = y)$	0.29	0.24	0.47

Отже, розподіл $\mathbb{P}(\xi, \eta)$:

-	$\eta = -1$	$\eta = 0$	$\eta = 2$
$\xi = 0$	$0.24 \cdot 0.29$	$0.24 \cdot 0.24$	$0.24 \cdot 0.47$
$\xi = 1$	$0.07 \cdot 0.29$	$0.07 \cdot 0.24$	$0.07 \cdot 0.47$
$\xi = 3$	$0.69 \cdot 0.29$	$0.69 \cdot 0.24$	$0.69 \cdot 0.47$

Або, остаточно, якщо порахувати:

-	$\eta = -1$	$\eta = 0$	$\eta = 2$
$\xi = 0$	0.0696	0.0576	0.1128
$\xi = 1$	0.0203	0.0168	0.0329
$\xi = 3$	0.2001	0.1656	0.3243