

11. ДЕЯКІ ДИСКРЕТНІ СТАТИСТИЧНІ РОЗПОДІЛИ

Дискретні розподіли часто зустрічаються на практиці. Найбільш важливими для фармацевтичного аналізу є біноміальний розподіл і розподіл Пуассона.

11.1. Біноміальний розподіл

11.1.1. Визначення

Якщо послідовно проводиться n незалежних випробувань (наприклад, кидання монети), у кожному з яких якась подія A (наприклад, «орел») відбувається з постійною ймовірністю p , то загальна кількість випробувань із наслідком a являє собою випадкову величину (позначимо її μ) з ймовірністю

$$P\{\mu = k \mid n, p\} = C_n^k \times p^k \times (1 - p)^{n-k}. \quad (11.1)$$

Оскільки ймовірності $p\{\mu = k \mid n, p\}$ являють собою члени розкладання бінома $(p + q)^n$ за $q = 1 - p$, то такий розподіл ймовірностей випадкової величини μ (для різних k) називають біноміальним розподілом із двома параметрами: n (число випробувань) і p (ймовірність події A).

Біноміальний коефіцієнт C_n^k — це кількість комбінацій із n по k , яку (кількість) розраховують за формулою:

$$C_n^k = \frac{(n-1) \times (n-2) \times \dots \times (n-k)}{1 \times 2 \times \dots \times k} = \frac{n!}{k!(n-k)!}. \quad (11.2)$$

Біноміальний розподіл (11.1-11.2) характеризується математичним сподіванням (генеральним середнім)

$$\bar{\mu} = n \times p, \quad (11.3)$$

дисперсією генеральної сукупності

$$\sigma^2 = n \times p \times (1 - p) \quad (11.4)$$

і стандартним відхилення генеральної сукупності

$$\sigma = \sqrt{n \times p \times (1 - p)}. \quad (11.5)$$

11.1.2. Застосування у фармацевтичному аналізі

Біноміальний розподіл застосовується у фармацевтичному аналізі під час вибіркового підтверджувального контролю якості продукції або якості результатів учасників тестування за альтернативною ознакою (тобто «відповідає» чи «не відповідає»).

Водночас можна виділити такі завдання.

11.1.2.1. Гранично допустима кількість дефектних одиниць під час підтвердження якості

Під час виробництва лікарських засобів обов'язковим є дотримання вимог Належної виробничої практики (GMP), що передбачає, зі свого боку, обов'язкову валідацію виробничого процесу і доведення того, що генеральна частка (імовірність) дефектних одиниць p не перевищує деяку, наперед встановлену, величину $\max p$. Тому на кінцевій стадії (і в мережі реалізації) під час контролю якості може бути застосовний підтверджувальний підхід (див. п. 2.3.3), у процесі якого, на підставі вибіркового контролю, перевіряють, чи не перевищує знайдена кількість одиниць (μ) дефектних одиниць (подія A) критичне (максимально при/допустиме) значення ($\max \mu$) для цієї вибірки обсягу n . Зі співвідношення (11.3) випливає:

$$\max \mu = n \times \max p \quad (11.6)$$

Під час такого контролю задають максимально допустиму генеральну частку дефектних одиниць $\max p$, яка була підтверджена на стадії валідації технологічного процесу, і ймовірність $P\{\mu \leq \max \mu \mid n, \max p\}$ (зазвичай це 95 % для одностороннього нормування), з якою фактична кількість дефектних одиниць μ не має перевищувати критичне значення $\max \mu$.

З огляду на співвідношення (11.1-11.2), ймовірність того, що фактична частка дефектних одиниць μ у вибірці обсягу n не перевищить $\max \mu$ одиниць, дорівнює:

$$P\{\mu \leq \max \mu \mid n, \max p\} = \sum_{k=1}^{k=\max \mu} P\{\mu = k \mid n, \max p\} = \sum_{k=1}^{k=\max \mu} C_n^k \times (\max p)^k \times (1 - \max p)^{n-k}. \quad (11.7)$$

Значення вагової (11.1) й інтегральної (11.7) функцій біноміального розподілу надаються в статистичних таблицях¹, а також у комп'ютерних програмах (наприклад, *Excel*). Для невеликих k вони можуть бути безпосередньо розраховані за співвідношеннями (11.1, 11.2, 11.7).

11.1.2.2. Довірчі межі для параметра p біноміального розподілу

Під час проведення n послідовних випробувань подія a (наприклад, незадовільний результат або брак) здійснювалася μ разів. Які довірчі межі для генерального значення параметра p (імовірності дефектних одиниць або незадовільного результату)? Таке завдання виникає, наприклад, під час визначення частки дефектних одиниць на стадії валідації технології виробництва або під час узагальненої оцінки результатів учасників тестування.

Довірчі межі для параметра p біноміального розподілу можна знайти в статистичних таблицях¹. Для невеликих вибірок під час узагальненої оцінки результатів тестування можна використовувати також Табл. 6.3 (див. розділ 6.4.4.2.2).

Застосування біноміального розподілу не викликає труднощів для невеликих вибірок n і кількості дефектів k . Однак для великих вибірок і кількості дефектів його застосування ускладнено. Наприклад, для $n = 1000$ і $k = 50$ під час розрахунків біноміального коефіцієнта C_n^k за співвідношенням (11.2) нам доведеться мати справу з числами порядку 10^{150} . Тому на практиці зазвичай використовують граничні форми біноміального розподілу.

Зокрема, за центральною граничною теоремою статистики, у разі обсягу вибірки $n \rightarrow \infty$ усі розподіли (включно з біноміальним) наближаються до нормального — головного — статистичного розподілу, застосування якого докладно розглянуто в попередніх розділах.

11.2. Розподіл Пуассона

У тому разі, коли можна вважати $n \times p = \text{const}$, а $n \rightarrow \infty$, біноміальний розподіл переходить у розподіл Пуассона.

Розподіл Пуассона є однопараметричним і значно простішим у застосуванні, ніж двопараметричний біноміальний. Тому він часто застосовується замість нього для великих вибірок під час розв'язання завдань, що описані вище в розділі 11.1.2.

Розподіл імовірностей випадкової величини μ , яка набуває цілих невід'ємних значень, називають розподілом Пуассона з одним параметром λ , якщо

$$P(\mu = k | \lambda) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad \lambda > 0, \quad k = 0, 1, 2, \dots \quad (11.8)$$

Цей розподіл характеризується математичним сподіванням (генеральним середнім)

$$\bar{\mu} = \lambda, \quad (11.9)$$

дисперсією генеральної сукупності (яка співпадає з математичним сподіванням)

$$\sigma^2 = \lambda \quad (11.10)$$

і стандартним відхиленням генеральної сукупності

$$\sigma = \sqrt{\lambda}. \quad (11.11)$$

За відомої ймовірності p події A , значення параметра λ розподілу Пуассона розраховують за співвідношеннями (11.2, 11.9).

Ймовірність того, що випадкова величина μ не перевищить максимальне значення $\max \mu$, дорівнює:

$$P(\mu \leq \max \mu | \lambda) = \sum_{k=0}^{k=\max \mu} \frac{\lambda^k}{k!} e^{-\lambda}. \quad \lambda > 0, \quad k = 0, 1, 2, \dots \quad (11.12)$$

Значення вагової (11.8) й інтегральної (11.12) функцій розподілу Пуассона представлені в статистичних таблицях¹, а також у комп'ютерних програмах (наприклад, *Excel*). Для невеликих значень k вони можуть бути безпосередньо розраховані за співвідношеннями (11.8-11.12).

Довірчі межі для генеральної кількості дефектної продукції (тобто параметра λ розподілу Пуассона) за результатами аналізу вибірки розміром n знаходять за Табл. 11.1.

Таблиця 11.1

Довірчі межі для параметра λ розподілу Пуассонаі

$2P - 1$	0.98		0.95		0.90	
P	0.99		0.975		0.95	
μ	λ_{min}	λ_{max}	λ_{min}	λ_{max}	λ_{min}	λ_{max}
0	0	4.61	0	3.69	0	3.00
1	0.0101	6.64	0.0253	5.57	0.0513	4.74
2	0.149	8.41	0.242	7.22	0.355	6.30
3	0.436	10.05	0.619	8.77	0.818	7.75
4	0.823	11.60	1.09	10.24	1.37	9.15
5	1.28	13.11	1.62	11.67	1.97	10.51

6	1.79	14.57	2.20	13.06	2.61	11.84
7	2.33	16.00	2.81	14.42	3.29	13.15
8	2.91	17.40	3.45	15.76	3.98	14.43
9	3.51	18.78	4.12	17.08	4.70	15.71
10	4.13	20.14	4.80	18.39	5.43	16.96
11	4.77	21.49	5.49	19.68	6.17	18.21
12	5.43	22.82	6.20	20.96	6.92	19.44
13	6.10	24.14	6.92	22.23	7.69	20.67
14	6.78	25.45	7.65	23.49	8.46	21.89
15	7.48	26.74	8.46	24.74	9.25	23.10
16	8.18	28.03	9.15	25.98	10.04	24.30
17	8.89	29.34	9.90	27.22	10.83	25.50
18	9.62	30.58	10.67	28.45	11.63	26.69
19	10.35	31.85	11.44	29.67	12.44	27.88
20	11.08	33.10	12.22	30.89	13.25	29.06
21	11.82	34.36	13.00	32.10	14.07	30.24
22	12.57	35.60	13.79	33.31	14.89	31.42
23	13.33	36.84	14.58	34.51	15.72	32.59
24	14.09	38.08	15.38	35.71	16.55	33.75
25	14.85	39.31	16.18	36.90	17.38	34.92
26	15.62	40.53	16.98	38.10	18.22	36.08
27	16.40	41.76	17.79	39.28	19.06	37.23
28	17.17	42.98	18.61	40.47	19.90	38.39
29	17.96	44.19	19.42	41.65	20.75	39.54
30	18.74	45.40	20.24	42.83	21.59	40.69
31	19.53	46.61	21.06	44.00	22.44	41.84
32	20.32	47.81	21.89	45.17	23.30	42.98
33	21.12	49.01	22.72	46.34	24.15	44.12
34	21.92	50.21	23.55	47.51	25.01	45.27
35	22.72	51.41	24.38	48.68	25.87	46.40
36	23.53	52.60	25.21	49.84	26.73	47.54
37	24.33	53.79	26.05	51.00	27.59	48.68
38	25.14	54.98	26.89	52.16	28.46	49.81
39	25.96	56.16	27.73	53.31	29.33	50.94
40	26.77	57.35	28.58	54.47	30.20	52.07
41	27.59	58.53	29.42	55.62	31.07	53.20
42	28.41	59.71	30.27	56.77	31.94	54.32
43	29.23	60.88	31.12	57.92	32.81	55.45

44	30.05	62.06	31.97	59.07	33.69	56.57
45	30.88	63.23	32.82	60.21	34.56	57.70
46	31.70	64.40	33.68	61.36	35.44	58.82
47	32.53	65.57	34.53	62.50	36.32	59.94
48	33.36	66.74	35.39	63.64	37.20	61.05
49	34.20	67.90	36.25	64.78	38.08	62.17
50	35.03	69.07	37.11	65.92	38.96	63.29

Застосування біноміального розподілу і розподілу Пуассона для розв'язання практичних завдань, які описані в розділі 11.1.2, наведено в *Прикладах 11.1 і 11.2.*

Приклад 11.1. Підтверджувальний вибірковий контроль якості ін'єкцій за механічними включеннями

На стадії валідації технологічного процесу отримання ін'єкційного розчину була показано, що генеральна частка дефектної за механічними включеннями продукції не перевищує 0.65 %, тобто $\max p = 0.0065$. У процесі технології препарат підлягає 100%-му контролю за механічними включеннями. Для вибіркового підтвердження якості за цим показником, відповідно до вимог одноступінчатого плану ISOⁱⁱ, для загального II рівня контролю, залежно від розміру партії (серії), беруться вибірки з відповідними критеріями допустимої наявності дефектної продукції, зазначені в Табл. 11.2.

Які ймовірності того, що продукція витримає ці вимоги?

Застосування біноміального розподілу

Підставляючи в формули (11.1, 11.2, 11.7) значення $\mu = 0, 1, 2$ і ураховуючи, що $c_n^0 = 1$, отримаємо:

- 1) $P\{\mu = 0 \mid 20, 0.0065\} = 1 \times 0.0065^0 \times (1 - 0.0065)^{20} = 0.878$,
- 2) $P\{\mu = 0 \mid 32, 0.0065\} = 1 \times 0.0065^0 \times (1 - 0.0065)^{32} = 0.812$,
- 3) $P\{\mu = 0 \mid 50, 0.0065\} = 1 \times 0.0065^0 \times (1 - 0.0065)^{50} = 0.722$,
- $P\{\mu = 1 \mid 50, 0.0065\} = (50 / 1) \times 0.0065 \times (1 - 0.0065)^{49} = 0.236$,
- $P\{\mu \leq 1 \mid 50, 0.0065\} = 0.722 + 0.236 = 0.958$ тощо.

Далі доцільно використовувати комп'ютерні програми, наприклад в *Excel*.

Результати розрахунків наведені в Табл. 11.2.

Застосування розподілу Пуассона

Використовуючи співвідношення (11.3, 11.6, 11.8, 11.9), отримаємо:

1) $n = 20$;

$$\lambda = 0.0065 \times 20 = 0.13,$$

$$P\{\mu = 0 \mid \lambda = 0.13\} = (0.13^0 / 0!) \times e^{-0.13} = 0.878.$$

2) $n = 32$;

$$\lambda = 0.0065 \times 32 = 0.208,$$

$$P\{\mu = 0 \mid \lambda = 0.208\} = (0.208^0 / 0!) \times e^{-0.208} = 0.812.$$

3) $n = 50$;

$$\lambda = 0.0065 \times 50 = 0.325,$$

$$P\{\mu = 0 \mid \lambda = 0.325\} = (0.325^0 / 0!) \times e^{-0.325} = 0.722,$$

$$P\{\mu = 1 \mid \lambda = 0.325\} = (0.325^1 / 1!) \times e^{-0.325} = 0.235,$$

$$P\{\mu \leq 1 \mid \lambda = 0.325\} = 0.722 + 0.235 = 0.957 \text{ тощо.}$$

Далі доцільно використовувати комп'ютерні програми, наприклад в *Excel*.

Результати розрахунків наведені в Табл. 11.2.

Таблиця 11.2

Імовірність того, що партію не забракують під час одноступінчатого плану
II загального рівня контролю відповідно до вимог ISOⁱⁱ

№	Розмір вибірки, одиниці	Максимально допустима кількість одиниць дефектної продукції μ	Імовірність того, що партію не забракують $100 \times P\{\mu \mid n, 0.0065\}$, %	
			Біноміальний розподіл	Розподіл Пуассона
1	20	0	87.8	87.8
2	32	0	81.2	81.2
3	50	1	95.8	95.7
4	80	1	90.4	90.4
5	125	2	95.1	95.1
6	200	3	95.7	95.7
7	315	5	98.2	98.2

8	500	7	98.2	98.2
---	-----	---	------	------

Як видно, значення, отримані з використанням розподілу Пуассона, співпадають із розрахунками за більш точним біноміальним розподілом. Тому розподіл Пуассона зазвичай використовують замість біноміального розподілу для малої кількості дефектних одиниць (зокрема, в ISOⁱⁱ).

Ураховуючи дискретність і невелику кількість дефектних одиниць для малих вибірок, можна зробити висновок, що критерії ISO орієнтовані на загальноприйнятий в аналітичній практиці рівень надійності 95 % для підтверджувального підходу (див. п. 2.3.3).

Приклад 11.2. Узагальнене оцінювання якості результатів учасників міжлабораторного тестування

Узагальнене оцінювання якості результатів учасників за допомогою біноміального розподілу є одним з елементів зовнішнього тестування лабораторій (див. розділ 8.9.2).

Під час проведення зовнішнього тестування $n = 60$ лабораторій $\mu = 7$ учасників показали незадовільні результати. Організатори тестування вважають, що генеральне значення частки лабораторій (p), які не витримують вимоги тестування, не має перевищувати 0.05 (або 5 %). Відповідно до співвідношень (11.3), це означає $\bar{\mu} = 3.0$. Чи можна вважати, що фактичне значення $\mu = 7$ статистично значуще відрізняється від $\bar{\mu} = 3.0$ за одностороннього коефіцієнта довіри (імовірності) $p = 0.95$?

Застосування біноміального розподілу

Для такого оцінювання зручно використовувати Табл. 6.3 (див. розділ 6.4.4.2.2). З неї випливає, що в нашому разі ($p = 0.05$, $n = 60$, $p = 0.95$) $\max \mu = 7.5$. Оскільки $\mu = 7 < \max \mu = 7.5$, то ми не можемо говорити про статистично значущу відмінність частки незадовільних результатів від 5 %.

Застосування розподілу Пуассона

Зі співвідношень (11.3) і (11.9) отримаємо $\mu = 0.05 \times 60 = 3.0$. Для $\mu = 3.0$ і $p = 0.95$ за Табл. 11.1 знаходимо $\lambda_{\max} = 7.75$ (а за біноміальним розподілом 7.5).

Оскільки $\mu = 7 < \max \mu = 7.75$, то ми не можемо говорити про статистично значущу відмінність частки незадовільних результатів від 5 %.

Як видно, розподіл Пуассона дає такі самі висновки, що й більш точний біноміальний розподіл.

У разі малих значень дефектів (k) розрахунки за розподілом Пуассона мало відрізняються від розрахунків за більш точним біноміальним розподілом. Тому він широко застосовується, наприклад, у документах ISOⁱⁱ.

ⁱ Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. Москва: Наука. 1983. 415 с.

ⁱⁱ ISO 2859-1: "Sampling procedures for inspection by attributes — Part 1: Sampling schemes indexed by acceptance quality limit (AQL) for lot-by-lot inspection".