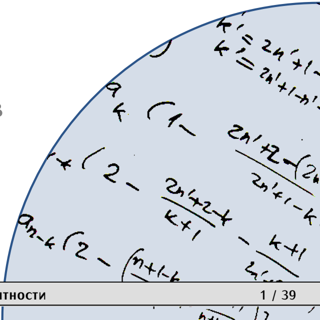
Введение в экономико-математическое моделирование

Лекция 13. Случайные события и вероятности – 2

Формула Байеса и повторение испытаний

канд. физ.-матем. наук, доцент Д. В. Чупраков

usr10381@vyatsu.ru



Структура лекции

1. Формула Байеса
   - Понятие полной вероятности
   - Формула Байеса
2. Повторение испытаний
   - Формула Бернулли
   - Локальная теорема Муавра — Лапласа
   - Интегральная теорема Муавра — Лапласа
   - Оценка отклонения частоты от вероятности
3. Резюме лекции и домашнее задание

Полная группа событий

События $H_1, H_2, \ldots, H_n$ образуют полную группу событий, если

1. события $H_1, H_2, \ldots, H_n$ попарно несовместны;
2. $H_1 + H_2 + \ldots + H_n = 1$

- События образуют полную группу, если в результате опыта происходит ровно одно из них.
- Все исходы опыта образуют полную группу событий.
- Событие A и противоположное к нему событие $\overline{A}$ образуют полную группу событий.

Полная вероятность

Следствием теорем о вероятностях является формула полной вероятности:

Теорема

Если

- $H_1, H_2, \ldots, H_n$ — *полная группа событий;*
- *известны условные вероятности* $p(A/H_i)$ *для всех* $i \in \{1, 2, \ldots, n\}$,

то

$$p(A) = \sum_{i=1}^{n} p(H_i)p(A/H_i) = p(H_1)p(A/H_1) + \ldots + p(H_n)p(A/H_n)$$

В магазин поступила новая продукция с трех предприятий. Процентный состав этой продукции следующий: 20% — продукция первого предприятия, 30% — продукция второго предприятия, 50% — продукция третьего предприятия; Вероятность брака продукции первого предприятия — 1%, второго предприятии - 0.5%, третьего — 2%. Найти вероятность того, что случайно купленная новая продукция окажется бракованной.

Формализуем задачу:

- Событие A — товар бракованный. Может наступить только с одним из несовместных событий H_1, H_2, H_3:

  H_1 — продукт 1 предприятия, $p(H_1) = 0.2$, $p(A/H_1) = 0.01$

  H_2 — продукт 2 предприятия, $p(H_2) = 0.3$, $p(A/H_2) = 0.005$

  H_3 — продукт 3 предприятия, $p(H_3) = 0.5$, $p(A/H_3) = 0.02$

Типовой пример. Полная вероятность II

- $p(H_1) + p(H_2) + p(H_3) = 1$ — это полная группа событий.
- Применим формулу полной вероятности

$$p(A) = p(H_1)p(A/H_1) + p(H_2)p(A/H_2) + p(H_3)p(A/H_3) =$$
$$= 0.2 \cdot 0.01 + 0.3 \cdot 0.005 + 0.5 \cdot 0.02 = 0.0135$$

- Итак, $p(A) = 0.0135$ или 1.35%.

Формула Байеса I

- Предположим, есть полная группа событий $H_1, H_2, \ldots, H_n$, вероятности которых известны и равны $p(H_1), p(H_2), \ldots, p(H_n)$.
- Назовем эти события гипотезами
- Пусть некоторое случайное событие A может произойти, только если осуществится одна из этих гипотез H_i. Причём известны вероятности $p(A/H_i)$
- На основании формулы полной вероятности

$$p(A) = p(H_1)p(A/H_1) + p(H_2)p(A/H_2) + \ldots + p(H_n)p(A/H_n)$$

Формула Байеса II

- Произведено испытание, в результате которого произошло событие A.
- Возникает вопрос, какова вероятность, что при этом осуществилась одна из гипотез: $H_1, H_2, \ldots, H_n$? То есть интересует $p(H_i/A)$.
- Посмотрим на вероятность появления события $H_i \cdot A$.

$$p(H_i \cdot A) = p(H_i)p(A/H_i) = p(A)p(H_i/A)$$

- Выразим $p(H_i/A)$:

$$p(H_i/A) = \frac{p(H_i)p(A/H_i)}{p(A)}$$

Формула Байеса III

Итак, имеет место теорема:

Теорема (Байеса)

Если

- $H_1, H_2, \dots, H_n$ — *полная группа событий;*
- *известны вероятности* $p(H_i)$ *для всех* $i \in \{1, 2, \dots, n\}$,
- *известны условные вероятности* $p(A/H_i)$ *для всех* $i \in \{1, 2, \dots, n\}$,

то

$$p(H_i/A) = \frac{p(H_i)p(A/H_i)}{\sum_{j=1}^{n} p(H_j)p(A/H_j)}$$

Типовой пример. Формула Байеса I

Допустим у меня есть комплект D&D дайсов. Я беру 3 кости: с 4, 6 и 8 гранями. Подбрасываю какую-то из них и говорю вам, что выпало 2. Какую кость я подбрасывал?

Типовой пример. Формула Байеса II

- Формализуем: событие A — выпала 2 могло произойти, в одном из трех случаев:

  H_1 — выбран D3, $p(H_1) = 1/3$, $p(A/H_1) = 1/3$

  H_2 — выбран D6, $p(H_2) = 1/3$, $p(A/H_2) = 1/6$

  H_3 — выбран D8, $p(H_3) = 1/3$, $p(A/H_3) = 1/8$
- Это априорные — доопытные вероятности
- Найдем вероятность A по формуле полной вероятности:

$$p(A) = p(H_1)p(A/H_1) + p(H_2)p(A/H_2) + p(H_3)p(A/H_3) =$$

$$= \frac{1}{3} \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{6} + \frac{1}{3} \cdot \frac{1}{8} = \frac{5}{24}.$$

- Применим теперь формулу Байеса:

$$p(H_1/A) = \frac{p(H_1)\cdot p(A/H_1)}{p(A)} = \frac{1/3\cdot 1/3}{5/24} = \frac{8}{15} \approx 0.53.$$
$$p(H_2/A) = \frac{p(H_2)\cdot p(A/H_2)}{p(A)} = \frac{1/3\cdot 1/6}{5/24} = \frac{4}{15} \approx 0.27.$$
$$p(H_3/A) = \frac{p(H_3)\cdot p(A/H_3)}{p(A)} = \frac{1/3\cdot 1/8}{5/24} = \frac{1}{5} \approx 0.20.$$

- Итак, я бросал
  - D3 с вероятностью 53%,
  - D6 — с вероятностью 27%,
  - D8 — с вероятностью 20%.

Знание результата существенно изменило оценку действий.

Пример на злобу дня I

Есть некоторое заболевание на К, которое диагностируют у 3% процентов населения страны. Экспресс-тест на вирусную РНК дает верные результаты в 94% случаев. Какова вероятность оказаться заболевшим, если тест показывает положительный результат?

- **Формализуем:** событие A — Тест показал положительный результат могло произойти, вместе со следующими событиями:

$$
\begin{array}{llll}
H_1 \text{ — пациент болен} & p(H_1) = 0.03, & p(A/H_1) = 0.94 \\
H_2 \text{ — пациент не болен} & p(H_2) = 0.97, & p(A/H_2) = 0.06
\end{array}
$$

Пример на злобу дня II

- Полная вероятность A — положительного результата теста равна:

$$p(A) = p(H_1)p(A/H_1) + p(H_2)p(A/H_2) = \\ = 0.03 \cdot 0.94 + 0.97 \cdot 0.06 = 0.0864.$$

- Применим теперь формулу Байеса чтобы определить, вероятность болезни у пациента:

$$p(H_1/A) = \frac{p(H_1) \cdot p(A/H_1)}{p(A)} = \frac{0.03 \cdot 0.94}{0.0864} \approx 0.326.$$

Итак, вероятность того, что пациент действительно болен не больше 33%.

Пример на злобу дня III

Тест проведен еще раз и снова получен положительный результат. Какова теперь вероятность оказаться заболевшим

- Вновь событие A — положительный результат теста, но априорные вероятности событий изменились:

  H_1 — пациент болен $\quad p(H_1) = 0.33, \quad p(A/H_1) = 0.94$

  H_2 — пациент не болен $\quad p(H_2) = 0.67, \quad p(A/H_2) = 0.06$

- Полная вероятность A — положительного результата теста равна:

$$p(A) = p(H_1)p(A/H_1) + p(H_2)p(A/H_2) =$$
$$= 0.33 \cdot 0.94 + 0.67 \cdot 0.06 = 0.3504.$$

- Применим теперь формулу Байеса чтобы определить, вероятность болезни у пациента:

$$p(H_1/A) = \frac{p(H_1)\cdot p(A/H_1)}{p(A)} = \frac{0.33\cdot 0.94}{0.3504} \approx 0.885$$

Итак, вероятность того, что пациент действительно болен после двухкратного положительного результата теста равна 89%.

Многократно применяемая формула Байеса — мощный инструмент принятия решений. Она позволяет последовательно уточнять вероятности гипотез о развитии управляемой системы в зависимости от события, возникающих при функционировании этой системы.

Допустим,

- Испытания независимы — появление события не влияет на вероятность появления события при следующих опытах.
- Вероятность $p = p(A)$ появления события A одинакова в каждом их опытов.
- Вероятность противоположного события $q = p(\overline{A}) = 1 - p$.
- Требуется найти вероятность $P_n(k)$ — наступления события A ровно k раз из n проведенных опытов.

В этом случае говорят, что поставлена задача повторения независимых испытаний.

Искомая вероятность обозначается $P_n(k)$.

Вероятность появления заданного числа событий

1. По теореме о произведении независимых событий:

$$P(\underbrace{A \cdot A}_{k}\underbrace{\overline{A} \cdot \overline{A}}_{n-k}) = p^k q^{n-k}.$$

2. Теперь заметим, что выбрать k испытаний, в которых произошло событие можно $C_n^k = \frac{n!}{k!(n-k)!}$ способами, при этом все эти события несовместны.
3. По теореме о сумме несовместных событий

$$P_n(k) = \underbrace{p^k q^{n-k} + \ldots + p^k q^{n-k}}_{C_n^k} = C_n^k p^k q^{n-k}.$$

Формула Бернулли

Формула Бернулли

Вероятность того, что при n *независимых испытаниях событие* A *осуществится ровно* k *раз равна*

$$P_n(k) = C_n^k p^k q^{n-k}.$$

Пример. Формула Бернулли

Неприменимость формулы Бернулли

Формула требует выполнения действий над громадными числами, поэтому пользоваться формулой Бернулли при больших значениях n достаточно сложно.

Локальная теорема Муавра — Лапласа

- Нельзя ли вычислить интересующую нас вероятность, не прибегая к формуле Бернулли?
- Да, можно, приближенно!
- Для этого используется функция Гаусса:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Локальная теорема Муавра — Лапласа

Если вероятность $p = p(A)$ в каждом испытании постоянна и $p \neq 1$, $p \neq 0$, то вероятность $P_n(k)$ того, что событие A появится в n испытаниях ровно k раз приближенно равна

$$P_n(k) \approx \frac{1}{\sqrt{npq}} \varphi(x), \quad \text{при} \quad x = \frac{k - np}{\sqrt{npq}}$$

Вычисление функции Гаусса

Для вычисления функции Гаусса

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

рекомендуется использовать:

- таблицы значений.
- функции электронных таблиц:
  - `=ФИ(x)` (MS Excel , LibreOffice Calc)
  - `=НОРМ.СТ.РАСП(x;0)` (MS Excel 2010, LibreOffice Calc)
  - `=PHI(x)` (Google spreadsheets)

Свойство четности функции Гаусса

$\varphi(-x) = \varphi(x)$

Типовой пример

Задача

Найти вероятность того, что при 400 испытаниях событие наступит ровно 104 раза, если вероятность его появления в каждом испытании равна $p = 0.2$.

- Исходные данные:
  $n = 400, \quad k = 104, \quad p = 0.2, \quad q = 1 - p = 0.8$
- n — большое, воспользуемся локальной теоремой Муавра — Лапласа.
- Вычислим $\sqrt{npq} = \sqrt{400 \cdot 0.2 \cdot 0.8} = \sqrt{64} = 8$
- Вычислим $x = \frac{k-np}{\sqrt{npq}} = \frac{104-400\cdot0.2}{8} = \frac{24}{8} = 3$
- Вычислим $\varphi(x) = \varphi(3) \approx 0.0044$ =ФИ(3)
- $P_{400}(104) \approx \frac{1}{\sqrt{npq}}\varphi(x) \approx \frac{1}{8}0.0044 \approx 0.00055$

Ответ: вероятность $P_{400}(104) \approx 0.00055$.

Проблема

Странная ситуация

При проведении рекламной кампании на проспекте был размещен рекламный стенд. Ежедневно его видят 5000 человек. Вероятность того, что человек среагирует на рекламу равна 0.1. Руководство интересует, вероятность того, что за месяц (30 дней) на рекламу среагируют ровно 15782 человека?

Проблема

Странная ситуация

При проведении рекламной кампании на проспекте был размещен рекламный стенд. Ежедневно его видят 5000 человек. Вероятность того, что человек среагирует на рекламу равна 0.1. Руководство интересует, вероятность того, что за месяц (30 дней) на рекламу среагируют ровно 15782 человека?

- Точное значение числа появлений события чаще всего не информативно.

Проблема

Странная ситуация

При проведении рекламной кампании на проспекте был размещен рекламный стенд. Ежедневно его видят 5000 человек. Вероятность того, что человек среагирует на рекламу равна 0.1. Руководство интересует, вероятность того, что за месяц (30 дней) на рекламу среагируют ровно 15782 человека?

- Точное значение числа появлений события чаще всего не информативно.
- При увеличении числа испытаний n вероятность того, что произойдет ровно k событий стремится к нулю.

Проблема

Странная ситуация

При проведении рекламной кампании на проспекте был размещен рекламный стенд. Ежедневно его видят 5000 человек. Вероятность того, что человек среагирует на рекламу равна 0.1. Руководство интересует, вероятность того, что за месяц (30 дней) на рекламу среагируют ровно 15782 человека?

- Точное значение числа появлений события чаще всего не информативно.
- При увеличении числа испытаний n вероятность того, что произойдет ровно k событий стремится к нулю.
- Интереснее какова вероятность, что число событий окажется в промежутке $k_1 \leqslant k \leqslant k_2$

Интегральная теорема Муавра — Лапласа

Если вероятность $p = p(A)$ *в каждом испытании постоянна и* $p \neq 1$, $p \neq 0$, *то вероятность* $P_n(k_1, k_2)$ *того, что событие* A *появится в* n *испытаниях от* k_1 *до* k_2 *раз приближенно равна определенному интегралу*

$$P_n(k_1, k_2) \approx \int_{x_1}^{x_2} \varphi(x)dx,$$

где

$$x_1 = \frac{k_1 - np}{\sqrt{npq}}, \quad x_2 = \frac{k_2 - np}{\sqrt{npq}}$$

Рассмотрим функцию Лапласа

$$\Phi(x) = \int_0^x \varphi(z)dz = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{z^2}{2}} dz$$

Интегральная теорема Муавра — Лапласа

Если вероятность $p = p(A)$ в каждом испытании постоянна и $p \neq 1$, $p \neq 0$, то вероятность $P_n(k_1, k_2)$ того, что событие A появится в n испытаниях от k_1 до k_2 раз приближенно равна определенному интегралу

$$P_n(k_1, k_2) \approx \Phi(x_2) - \Phi(x_1)$$

где $x_1 = \dfrac{k_1 - np}{\sqrt{npq}}, \quad x_2 = \dfrac{k_2 - np}{\sqrt{npq}}$

Функция Лапласа (интеграл Гаусса)

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-\frac{z^2}{2}}\, dz$$

не берущийся интеграл.

При решении задач пользуются специальной таблицей значений или функциями электронных таблиц

- В таблице присутствуют значения $\Phi(x)$ для $0 \leqslant x \leqslant 5$.
  - если $x < 0$, то $\Phi(x) = -\Phi(-x)$ Функция $\Phi(x)$ — нечетная;
  - если $x > 50$, то $\Phi(x) = 0.5$;
- функции электронных таблиц:
  - `=ГАУСС(x)` (MS Excel , LibreOffice Calc)
  - `=GAUSS(x)` (Google Sheets)

Поставим перед собой задачу:

Найти вероятность того, что отклонение относительной частоты $\frac{k}{n}$ от постоянной вероятности p по абсолютной величине не превышает заданного числа $\Delta > 0$: $p\left(\left|\frac{k}{n} - p\right| \leqslant \Delta\right)$

- Преобразуем неравенство:

$$\left|\frac{k}{n} - p\right| \leqslant \Delta$$

$$-\Delta \leqslant \frac{k}{n} - p \leqslant \Delta$$

$$-\Delta n \leqslant k - pn \leqslant \Delta n$$

$$pn - \Delta n \leqslant k \leqslant pn + \Delta n$$

- Оценим вероятность

$$p\left(\left|\frac{k}{n}-p\right| \leqslant \Delta\right) = p\left(pn-\Delta n \leqslant k \leqslant pn+\Delta n\right) = \\ = P_n(pn-\Delta n, pn+\Delta n)$$

- Применим интегральную теорему Муавра—Лаплпса. Для этого вычислим

$$x_1 = \frac{(k_1-pn)}{\sqrt{npq}} = \frac{(pn-\Delta n-pn)}{\sqrt{npq}} = -\Delta\sqrt{\frac{n}{pq}}$$

$$x_2 = \Delta\sqrt{\frac{n}{pq}}$$

- Наконец, вычислим искомую вероятность

$$
\begin{aligned}
&p\left(\left|\frac{k}{n}-p\right| \leqslant \Delta\right) \approx \\
&\qquad \approx P_n(pn-\Delta n, pn+\Delta n) = \Phi(x_2)-\Phi(x_1) = \\
&\qquad = \Phi\left(\Delta\sqrt{\frac{n}{pq}}\right) - \Phi\left(-\Delta\sqrt{\frac{n}{pq}}\right) = \\
&\qquad = \Phi\left(\Delta\sqrt{\frac{n}{pq}}\right) + \Phi\left(\Delta\sqrt{\frac{n}{pq}}\right) = \\
&\qquad\qquad = 2\Phi\left(\Delta\sqrt{\frac{n}{pq}}\right)
\end{aligned}
$$

Итак,

Теорема

Вероятность того, что отклонение относительной частоты $\nu = \frac{k}{n}$ *от постоянной вероятности* p *по абсолютной величине не превышает заданного числа* $\Delta > 0$, *приближенно равна значению удвоенной функции Лапласа* $\Phi(x)$ *при* $x = \Delta\sqrt{\frac{n}{pq}}$:

$$p\left(\left|\nu - p\right| \leqslant \Delta\right) \approx 2\Phi\left(\Delta\sqrt{\frac{n}{pq}}\right)$$

Эта теорема позволяет показать, насколько точна формула статистической вероятности.

Типовая задача

Вероятность появления события в каждом из 10000 независимых испытаний $p = 0.75$. Найти вероятность того, что относительная частота появления события отклонится от его вероятности по абсциссе величине не более чем на 0.001.

- $n = 10000$, $p = 0.75$, $q = 0.25$, $\Delta = 0.001$
- $p(|\nu - p|) \approx 2\Phi\left(\Delta\sqrt{\frac{n}{pq}}\right) = 2\Phi\left(0.001\sqrt{\frac{10000}{0.75 \cdot 0.25}}\right) =$
  $= 2\Phi(0.23) \approx 2 \cdot 0.09095 \approx 0.182$

Итак, вероятность отклонения частоты появления события от его вероятности не более чем на 0.001 равна 0.182 или 18.2%

Типовая задача

Вероятность появления события в каждом из независимых испытаний равна $p = 0.2$. Найти, какое отклонение относительной частоты появления события от его вероятности можно ожидать с вероятностью 0.9128 при 5000 испытаниях.

- $n = 5000$, $p = 0.2$, $q = 0.8$, $P = p(|\nu - p| \leqslant \Delta) = 0.9128$.
- $2\Phi\left(\Delta\sqrt{\frac{n}{pq}}\right) \approx p(|\nu - p|) = 0.9128$
- $\Phi\left(\Delta\sqrt{\frac{n}{pq}}\right) \approx = \frac{0.9128}{2} = 0.4564$
- Вычислим обратную функцию Φ по таблице. Найдем на пересечении какой строки и столбца расположено число 0.4564: $\Phi(1.82) = 0.4562$
- $\Delta\sqrt{\frac{n}{pq}} = 1.82$ $\qquad \Delta = 1.82\sqrt{\frac{pq}{n}} = 1.82\sqrt{\frac{0.2 \cdot 0.8}{5000}} \approx 0.01029$

Итак, $\Delta \approx 0.0103$.

Резюме

После проработки лекции вы должны знать:

- Формулу полную вероятности события;
- Формулу Байеса и ее применение;
- Формулу Бернулли решения задачи повторения независимых испытаний.
- Локальную и интегральную теоремы Муавра — Лапласа.

Задание

Для завершения лекции вам необходимо подготовить конспект, в который должны войти:

1. Формула Байеса + пример;
2. Задача повторения испытаний. Формула Бернулли (без доказательства) + пример
3. Формула Пуассона с указанием границ применения. (Учебник Кремера. с. 71) + пример
4. Локальная теорема Муавра — Лапласа с указанием границ применения + пример.
5. Интегральная теорема Муавра — Лапласа + пример.

Источники информации

- Формула Байеса Кремер Н. Ш. Теория вероятностей и математическая статистика . §1.11, с. 51–56.
- Повторение испытаний Кремер Н. Ш. Теория вероятностей и математическая статистика . Глава 2, с. 68–89.
- Таблица значений функций Гаусса
- Таблица значений функций Лапласа
- Все материалы по курсу здесь: https://cloud.mail.ru/public/48BX/47oESuaQQ