Осенний коллоквиум курса «Теория вероятностей»

ФКН НИУ ВШЭ, 2-й курс ОП ПМИ, 2-й модуль, 2016 учебный год

Билет 5

Схема Бернулли. Теорема Муавра-Лапласа (формулировка, доказательство только локальной теоремы и только для симметричного случая)

Схема Бернулли

Проводится N опытов, в каждом из которых может произойти определенное событие ("успех") с вероятностью p или не произойти ("неуспех") с вероятностью $q = 1 - p$.
Например, рассмотрим следующий эксперимент: N раз бросается монета с вероятностью выпадения орла (успеха) p, причем результат одного бросания не влияет на результат других бросаний.
Нас интересует число успехов, то есть в примере это число выпадения орла.
Можно считать, что множество элементарных исходов состоит из последовательностей длины N из нулей и единиц, где 1 соответствует успеху. Каждому исходу w с k единицами сопоставляем вероятность $p^k q^{N-k}$.

Определение 1. Построенное вероятностное пространство называют *схемой Бернулли*.

Утверждение.
Вероятность того, что в исходе ровно k единиц равна $C_N^k p^k q^{N-k}$.

Доказательство. P(ровно k единиц) равна сумме вероятностей вида $p^k q^{N-k}$ по всем исходам, в которых ровно k единиц, а это значит, что

$$P(\text{ровно } k \text{ единиц}) = C_N^k p^k q^{N-k}.$$

□

Утверждение.

$$\sum_{k=0}^{N} C_N^k p^k q^{N-k} = 1$$

(Если это выполнено, то наше вероятностное пространство корректно определено)

Доказательство.

$$1 = 1^n = (p+q)^n \overset{\text{по биному Ньютона}}{=} \sum_{k=0}^{N} C_N^k p^k q^{N-k}.$$

□

Определение 2. Набор вероятностей $P_{N,0}, P_{N,1}, \ldots, P_{N,N}$ называется *распределением Бернулли.*

Замечание 1. *(Формула Стирлинга)*

$$n! = \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n+\frac{\varepsilon_n}{12n}}, \text{ где } \varepsilon_n \in (0,1).$$

Теорема Муавра-Лапласа
(формулировка, доказательство только локальной теоремы и только для симметричного случая)

Определение 3. Функция
$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$
называется *функцией Гаусса.*

Теорема 1. *(Теорема Муавра-Лапласа)*
У нас есть схема Бернулли:

$$N \text{ подбрасываний}$$
$$k \text{ - число успехов}$$
$$p \text{ - вероятность успеха}$$
$$q = 1 - p$$

$$X_{N,k} = \frac{k - Np}{\sqrt{Npq}}$$

1 случай. (Локальная теорема Муавра-Лапласа)
Если k выбирается так, что
$$|X_{N,k}| \leqslant C, \text{ где } C \text{ не зависит от } N,$$
то $P_{N,k} = \frac{1}{\sqrt{Npq}} \phi(X_{N,k}) \cdot (1 + O(\frac{1}{\sqrt{N}}))$.

2 случай. (Интегральная теорема Муавра-Лапласа)
Для любых чисел $a < b$ имеем
$$P\left(a \leqslant \frac{k - Np}{\sqrt{Npq}} \leqslant b\right) \xrightarrow[n \to \infty]{} \int\limits_a^b \phi(x) dx.$$

Здесь в левой части написана вероятность того, что число единиц k лежит в диапазоне от $Np + a\sqrt{Npq}$ до $Np + b\sqrt{Npq}$.
Отметим, что во втором случае разница между вероятностью и интегралом оценивается через $\frac{p^2+q^2}{\sqrt{Npq}}$ и эта оценка точна. Следовательно, если p близко к нулю или к единице, то вероятность плохо приближается интегралом от ϕ.

Доказательство. (Доказательство локальной теоремы только для симметричного случая)
Пусть $N = 2n$. Найдем вероятность того, что в последовательности длины N ровно n единиц (ровно половина):
$$P_{2n,n} = C_{2n}^n \cdot \frac{1}{2^{2n}} = \frac{(2n)!}{(n!)^2 \cdot 2^{2n}}$$
Раскрываем по формуле Стирлинга и получаем следующее:
$$\frac{\sqrt{2\pi} \cdot (2n)^{2n+\frac{1}{2}} \cdot e^{-2n+O(\frac{1}{2n})}}{(\sqrt{2\pi})^2 \cdot n^{2n+1} \cdot e^{-2n+O(\frac{1}{n})} \cdot 2^{2n}} = \frac{1}{\sqrt{\pi n}} e^{O(\frac{1}{n})} = \frac{1}{\sqrt{\pi n}} \left(1 + O\left(\frac{1}{n}\right)\right)$$
Таким образом, $P_{2n,n} \sim \frac{1}{\sqrt{\pi n}} = \frac{1}{\sqrt{2n \cdot \frac{1}{4}}} \cdot \frac{1}{\sqrt{2\pi}} e^{\frac{(n-n)^2}{(2n \cdot \frac{1}{4})}}$ (что и требовалось).
Найдем теперь $P_{2n,k}$.
Заметим, что
$$P_{2n,n+a} = P_{2n,n-a} \text{ (следует из равенства } C_N^k = C_N^{N-k}).$$
Тогда будет достаточно найти одну из этих вероятностей. Найдем $P_{2n,n+a}$.
$$\frac{P_{2n,n+a}}{P_{2n,n}} = \frac{C_{2n}^{n+a} \cdot (\frac{1}{2})^{2n}}{C_{2n}^n \cdot (\frac{1}{2})^{2n}} = \frac{C_{2n}^{n+a}}{C_{2n}^n} = \frac{n! \cdot n!}{(n+a)!(n-a)!} =$$
$$= \frac{(n-a+1) \cdot (n-a+2) \cdot \ldots \cdot (n-1) \cdot n}{(n+1) \cdot (n+2) \cdot \ldots \cdot (n+a)} =$$
$$= \frac{(1 - \frac{a-1}{n}) \cdot (1 - \frac{a-2}{n}) \cdot \ldots \cdot (1 - \frac{1}{n}) \cdot 1}{(1 + \frac{1}{n}) \cdot (1 + \frac{2}{n}) \cdot \ldots \cdot (1 + \frac{a}{n})} =$$
$$= e^{ln(1-\frac{a-1}{n})+ln(1-\frac{a-2}{n})+\ldots+ln(1-\frac{1}{n})-ln(1+\frac{1}{n})-\ldots-ln(1+\frac{a}{n})}$$

Вспомним, что $ln(1+x) = x + O(x^2)$ при $-1 < x \leqslant 1$. Тогда найдем оценку степени, которую мы нашли:

$$ln\left(1-\frac{a-1}{n}\right)+ln\left(1-\frac{a-2}{n}\right)+\ldots+ln\left(1-\frac{1}{n}\right)-ln\left(1+\frac{1}{n}\right)-\ldots-ln\left(1+\frac{a}{n}\right)=$$

$$=\left(-\frac{a-1}{n}-\frac{a-2}{n}-\ldots-\frac{1}{n}\right)+\left(-\frac{1}{n}-\frac{2}{n}-\ldots-\frac{a-1}{n}-\frac{a}{n}\right)+O\left(\frac{a^2}{n^2}\right)=$$

$$=-\frac{2(1+2+\ldots+(a-1))+a}{n}+O\left(\frac{a^2}{n^2}\right)=-\frac{a(a-1)+a}{n}+O\left(\frac{a^2}{n^2}\right)=-\frac{a^2}{n}+O\left(\frac{a^2}{n^2}\right)$$

Итак,

$$P_{2n,n+a}=P_{2n,n}\cdot e^{-\frac{a^2}{n}+O\left(\frac{a^2}{n^2}\right)}=\frac{1}{\sqrt{\pi n}}\cdot e^{-\frac{a^2}{n}+O\left(\frac{a^2}{n^2}\right)}$$

Пусть $|a| \leqslant c \cdot \sqrt{n}$.

$$P_{2n,n+a}=\frac{1}{\sqrt{\pi n}}e^{\frac{-a^2}{n}}\left(1+O\left(\frac{1}{\sqrt{N}}\right)\right)$$

$\square$