Probability Theory & Statistics

Innopolis University, BS-I,II

Spring Semester 2016-17

Lecturer: Nikolay Shilov

Part I

POISSON LAW OF RARE EVENTS

Poisson Limit Theorem: statement

- Assume we are given an infinite series of trials $X_1, \ldots X_n, \ldots$ where
  - X_n= binomial(n, p_n)
  - $p_n = \mu/n$ (where μ is a constant).
- For every n>0 let $P_n(m)$ be the probability of m positive outcomes in the n^{th} trial; then this sequence converges to $\mu^m e^{-\mu}/m!$

Poisson Limit Theorem: proof

- $P_n(m) = C_n^m p_n^m (1 - p_n)^{n-m} =$
$$= C_n^m \left(\frac{\mu}{n}\right)^m \left(1 - \frac{\mu}{n}\right)^{n-m}$$

- $C_n^m \left(\frac{\mu}{n}\right)^m = \frac{n(n-1)\ldots\big(n-(m-1)\big)}{m!n^m}\mu^m \underset{n\to\infty}{\longrightarrow} \frac{\mu^m}{m!}$

Poisson Limit Theorem: proof

- $\lim_{x\to 0}(1-x)^{1/x} = e^{-1}$
- $\left(1-\frac{\mu}{n}\right)^{n-m} = \left[\left(1-\frac{\mu}{n}\right)^{n/\mu}\right]\mu\left(1-\frac{\mu}{n}\right)^{-m}$

$$\xrightarrow[n\to\infty]{} e^{-\mu}$$

- $\lim_{n\to\infty} P_n(m) = \lim_{n\to\infty} C_n^m p_n^m (1-p_n)^{n-m}$

$$= \frac{\mu^m}{m!}e^{-\mu}$$

Pragmatics

- Number of trials n is *big*,
- the probability p is *small*,
- μ= np is *neither big nor small*.

“Palindrome Ticket” example

- A *palindrome* is a word, number, etc., which reads the same backward as forward, such as *madam* or *racecar*.
- Usually municipal transport tickets in Russia have numeric numbers represented by six decimal digits.
- A six-digit number is palindrome iff it looks like *abccba*, where *a*, *b*, and *c* are decimal digits.

“Palindrome Ticket” example (cont.)

- What is probability to buy *exactly* 2 palindrome tickets travelling by public transport 100 times?
- Exercise: compute the probability directly.

"Palindrome Ticket" example (cont.)

- Applying Poisson Theorem:
  - the probability of a palindrome $p=0.001$ is small;
  - the number of rides (i.e. Bernoulli trials) $n=100$ is big;
  - product $\mu= np =0.1$ is neither big nor small;
  - $P_{100}(2) \approx 0.0045$.

“Lucky Ticket” exercise

(https://ru.wikipedia.org/wiki/Счастливый_билет)

- Lucky ticket in Russia is a municipal transport ticket where the sum of the first 3 digits of its number equals to the sum of the last 3 digits of the number:

  abcdef is lucky, if $a+b+c = d+e+f$.

- What is probability to buy *exactly* 2 lucky tickets travelling by public transport 100 times? Compute the probability directly and using Poisson Theorem.

Part II

DE MOIVRE-LAPLACE THEOREMS

Local Theorem: statement

- Assume we are given an infinite series of trials $X_1, \ldots X_n, \ldots$ where X_n= binomial(n, p) and 0<p<1.
- For every n>0 let $P_n(m)$ be the probability of m positive outcomes in the n^{th} trial; then

$$P_n(m) = \frac{\phi_0(t_n)}{\sqrt{npq}}(1+\alpha_n)$$

where $\phi_0(t_n) = \frac{1}{\sqrt{2\pi}}e^{-t_n^2/2}$, $t_n = \frac{m-np}{\sqrt{npq}}$

And $|\alpha_n| < \frac{c}{\sqrt{n}}$.

In simple words

- For big, large and huge n

$$P_{B(n,p)}(m) \approx \frac{\phi_0(t_n)}{\sqrt{npq}}$$

Innopolis Shuttle example

- There are 2,000 residents in Innopolis, I am acquainted with 400 of them.
- In April I plan a business and will take shuttle to railway-station and then back.
- In 2 trips in the shuttle I’ll meet 100 Innopolis residents.
- What is probability that I’ll meet (exactly) 20 people with whom I am acquainted?

Innopolis Shuttle example (cont.)

- Here:
- $p=0.2$, $q=0.8$, $n=100$, $m=20$;
- t (i.e. t_n) $= (m-np)/(npq)^{1/2} =$

  $= (20 - 100*0.2)/(100*0.2*0.8)^{1/2} = 0$;
- $\varphi_0(t) = \varphi_0(0) = 1/(2\pi)^{1/2} \approx 0.40$;
- $P_{100}(20) \approx 0.40/4 = 0.10$.

Innopolis Shuttle example: discussion

- Why the probability $P_{100}(20)$ is 0.1 but not p=0.2?
- Simply because of the question was about *exact* number 20, not about *approximately* 20.
- If to interpret term "*approximately* 20" as "*from* 15 *to* 25" then $P_{100}(15)$ + … $P_{100}(25)$ is (almost) 1.

Toward Theorem

- $P_n(m_1, m_2) = P_{B(n,p)}([m_1..m_2]) = \sum_{m=m_1}^{m=m_2} P_{B(n,p)}(m)$
- $P_n(m_1, m_2) \approx \sum_{m=m_1}^{m=m_2} \frac{\phi_0(x_m)}{\sqrt{npq}}$
- $x_m = \frac{m-np}{\sqrt{npq}}, \ (m_1 \leq m \leq m_2)$
- $\Delta x_m = x_{m+1} - x_m = \frac{(m+1)-np}{\sqrt{npq}} - \frac{m-np}{\sqrt{npq}} = \frac{1}{\sqrt{npq}}$

Toward Theorem (cont.)

- $P_n(m_1, m_2) \approx \sum_{m=m_1}^{m=m_2} \phi_0(x_m)\Delta x_m \approx$

  $$\approx \int_{x_{m_1}}^{x_{m_2}} \phi_0(x)dx = \frac{1}{\sqrt{2\pi}} \int_{x_{m_1}}^{x_{m_2}} e^{-x^2/2}dx$$

- $x_{m_1} = \frac{m_1 - np}{\sqrt{npq}}$

- $x_{m_2} = \frac{m_2 - np}{\sqrt{npq}}$

Функция Лапласа/
/Gauss error function

$$\Phi_0(x) = \int_0^x \phi_0(x)dx = \frac{1}{\sqrt{2\pi}}\int_0^x e^{-x^2/2}dx$$

x	0	0.5	1	1.5	2	2.5	3
Φ	0	0.192	0.341	0.433	0.477	0.494	0.499

- Problem: examine properties of the error function.
- Caution: Please be aware that constant factor may be different.

De Moivre-Laplace theorem: statement

- Assume we are given an infinite series of trials $X_1, \ldots X_n, \ldots$ where X_n= binomial(n, p) and $0<p<1$.
- For every n>0 let $P_n(m)$ be the probability of m positive outcomes in the n^{th} trial and $P_n(m_1,m_2) = P_n(m_1 \leq m \leq m_2)$; then

$$P_n(m_1, m_2) \approx \Phi(x_{m_2}) - \Phi(x_{m_1})$$

$$x_{m_1} = \frac{m_1 - np}{\sqrt{npq}} \quad x_{m_2} = \frac{m_2 - np}{\sqrt{npq}}$$

Exercise: back to shuttle example

- The last sentence from slide 16 reads: if to interpret term "*approximately* 20" as "*from* 15 *to* 25" then $P_{100}(15) + \ldots P_{100}(25)$ is (almost) 1.
- Validate this claim using de Moivre-Laplace theorem.