N.I. Lobachevsky State University
of Nizhni Novgorod

Probability theory and mathematical statistics:

Limit theorems for Bernoulli trials

Associate Professor
A.V. Zorine

Example. There are two vessels A and B, each with a volume of 1 dm^3. Each contains $2{,}7 \times 10^{23}$ molecules of gas. The vessels are brought into contact so that there is a free exchange of molecules between them. What is the probability that after 24 hours one of the vessels will have at least *one ten-thousand millionth* (i.e. *one ten-billionth*) part more molecules than the other?

Example. There are two vessels A and B, each with a volume of 1 dm^3. Each contains $2{,}7 \times 10^{23}$ molecules of gas. The vessels are brought into contact so that there is a free exchange of molecules between them. What is the probability that after 24 hours one of the vessels will have at least *one ten-thousand millionth* (i.e. *one ten-billionth*) part more molecules than the other?

Solution. For each molecule, the probability of being in one or the other vessel 24 hours later is the same and is $1/2$. Thus, it is as if $(5{,}4 \times 10^{23})$ trials were performed, for each of which the probability of being in vessel A is equal to $1/2$. Let μ be the number of molecules that go to vessel A and, hence, $(5{,}4 \times 10^{23} - \mu)$ is the number of molecules that go to vessel B.

Solution. For each molecule, the probability of being in one or the other vessel 24 hours later is the same and is $1/2$. Thus, it is as if $(5{,}4 \times 10^{23})$ trials were performed, for each of which the probability of being in vessel A is equal to $1/2$. Let μ be the number of molecules that go to vessel A and, hence, $(5{,}4 \times 10^{23} - \mu)$ is the number of molecules that go to vessel B. We have to determine the probability that

$$\left|\mu - (5{,}4 \times 10^{23} - \mu)\right| \geqslant \frac{5{,}4 \times 10^{23}}{10^{10}},$$

in other words, we must find the probability

$$P = \mathrm{P}\left(\left\{\left|\mu - 2{,}7 \times 10^{23}\right| \geqslant 2{,}7 \times 10^{13}\right\}\right)$$

By the addition theorem,

$$P = \sum_{|k - 2{,}7\times 10^{23}| \geqslant 2{,}7\times 10^{13}} b(k; 5{,}4 \times 10^{23}, 1/2).$$

We are interested in good approximate formulae for

$$b(k;n,p), \quad \sum_{k=a}^{b} b(k;n,p).$$

An approximation to $b(k;n,p)$ is called *a local approximation*. An approximation to $\sum_{k=a}^{b} b(k;n,p)$ is called *a integral approximation*.

For two sequences, $\{a_n; n = 0, 1, \ldots\}$ and $\{b_n; n = 0, 1, \ldots\}$,

$$a_n \sim b_n$$

means that

$$\frac{a_n}{b_n} \to 1 \quad \text{as } n \to \infty$$

Stirling's formula

For any integer $n \geqslant 1$,

$$n! = \sqrt{2\pi n}\, n^n e^{-n+\theta(n)}, \quad \frac{1}{12n+1} < \theta(n) < \frac{1}{12n}.$$

That is,

$$n! \sim \sqrt{2\pi n}\, n^n e^{-n}.$$

Put

$$H(x) = x \ln \frac{x}{p} + (1 - x) \ln \frac{1 - x}{1 - p}.$$

Theorem

$$b(k; n, p) = \frac{1}{\sqrt{2\pi n p^*(1 - p^*)}} e^{-nH(p^*) + \theta(k,n)}, \quad p^* = \frac{k}{n}$$

where

$$|\theta(k, n)| = |\theta(n) - \theta(k) - \theta(n - k)| < \frac{1}{12 n p^*(1 - p^*)}$$

Proof is streightforward. Apply Stirling's formula to Bernoulli formula for $b(k; n, p)$.

Proof.

$$
\begin{aligned}
C_n^k p^k (1-p)^{n-k} &= \frac{n!}{k!(n-k)!} p^k (n-k)^{n-k} \\
&= \frac{\sqrt{2\pi n}\, n^n e^{-n+\theta(n)} p^k (1-p)^{n-k}}{\sqrt{2\pi k}\, k^k e^{-k+\theta(k)} \sqrt{2\pi(n-k)}\, (n-k)^{n-k} e^{-(n-k)+\theta(n-k)}} \\
&= \sqrt{\frac{n}{2\pi k(n-k)}}\, \frac{n^n}{k^k (n-k)^{n-k}}\, p^k (1-p)^{n-k} e^{\theta(k,n)} \\
&= \frac{1}{\sqrt{2\pi n p^*(1-p^*)}}\, e^{k \ln \frac{p}{p^*} + (n-k) \ln \frac{1-p}{1-p^*} + \theta(k,n)} \\
&= \frac{1}{\sqrt{2\pi n p^*(1-p^*)}}\, e^{-nH(p^*)+\theta(k,n)}
\end{aligned}
$$

Example. Compute $b(3091; 9000, 1/3)$.

$$\begin{aligned} p &= 1/3 \approx 0{,}3333333, \\ p^* &= 3091/9000 \approx 0{,}3434444, \\ nH(p^*) &= 2{,}0600139, \\ (\sqrt{2\pi np^*(1-p^*)})^{-1} &= 0{,}0088557, \\ (12np^*(1-p^*))^{-1} &= 0{,}0000411. \end{aligned}$$

Now,

$$0{,}0088557 \cdot e^{-2{,}060055} < b(3091; 9000, 1/3) < 0{,}0088557 \cdot e^{-2{,}0599728},$$

$$0{,}0011286 < b(3091; 9000, 1/3) < 0{,}0011287.$$

In fact, $b(3091; 9000, 1/3) = 0{,}001128642...$

Observe that

$$H'(x) = \ln\frac{x}{p} - \ln\frac{1-x}{1-p}, \quad H''(x) = \frac{1}{x} + \frac{1}{1-x},$$
$$H(p) = 0, \quad H'(p) = 0,$$

and as $p^* - p \to 0$

$$H(p^*) = \frac{1}{2}\left(\frac{1}{p} + \frac{1}{1-p}\right)(p^* - p)^2 + O(|p^* - p)^3|).$$

Now, if $p^* \sim p$ and $n(p^* - p)^3 \to 0$ as $n \to \infty$ then

$$b(k; n, p) \sim \frac{1}{\sqrt{2\pi np(1-p)}} \exp\left\{-\frac{n}{2p(1-p)}(p^* - p)^2\right\}.$$

Define

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

Local theorem of De Moivre – Laplace

For $0 < p < 1$ and k, $|k - np| \cdot (np(1-p))^{-2/3} \to 0$ as $n \to \infty$,

$$b(k; n, p) \sim \frac{1}{\sqrt{np(1-p)}} \varphi(x_k), \quad x_k = \frac{k - np}{\sqrt{np(1-p)}}$$

Define

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

Local theorem of De Moivre – Laplace

For $0 < p < 1$ and k, $|k - np| \cdot (np(1-p))^{-2/3} \to 0$ as $n \to \infty$,

$$b(k; n, p) \sim \frac{1}{\sqrt{np(1-p)}} \varphi(x_k), \quad x_k = \frac{k - np}{\sqrt{np(1-p)}}$$

$\underline{n = 9000, p = 1/3:}$
$x_{3091} = 2{,}0348219$, $\varphi(x_{3091}) = 0{,}0503286$,
$(\sqrt{np(1-p)})^{-1} = 0{,}0223607$,
$\underline{b(3091; 9000, 1/3) \approx 0{,}0011253}$.

Let $p = 0{,}2$. Consider cases $n = 4$, $n = 25$.

k	x_k	$b(k;n,p)$	$\sqrt{np(1-p)}b(k;n,p)$	$\varphi(x_k)$
0	$-1{,}0000$	0,4096	0,3277	0,2420
1	0,2500	0,4096	0,3277	0,3867
2	1,5000	0,1536	0,1229	0,1295
3	2,7500	0,0256	0,0205	0,0091
4	4,0000	0,0016	0,0013	0,0001

k	x_k	$b(k;n,p)$	$\sqrt{np(1-p)}b(k;n,p)$	$\varphi(x_k)$
0	$-2{,}5000$	0,0038	0,0076	0,0175
1	$-2{,}0000$	0,0236	0,0472	0,0540
2	$-1{,}5000$	0,0708	0,1417	0,1295
3	$-1{,}0000$	0,1358	0,2715	0,2420
4	$-{,}5000$	0,1867	0,3734	0,3521
5	0,0000	0,1960	0,3920	0,3989
6	0,5000	0,1633	0,3267	0,3521
7	1,0000	0,1108	0,2217	0,2420
8	1,5000	0,0623	0,1247	0,1295
9	2,0000	0,0294	0,0589	0,0540
10	2,5000	0,0118	0,0236	0,0175
11	3,0000	0,0040	0,0080	0,0044
12	3,5000	0,0012	0,0023	0,0009
13	4,0000	0,0003	0,0006	0,0001
14	4,5000	0,0001	0,0001	0,00002
15	5,0000	0,00001	0,00002	0,00000
> 15	$> 5{,}0000$	0,00000	0,00000	0,00000

Let $p_1, p_2, \ldots, p_r$ be the probabilities of occurence, respectively of events $A_1, A_2, \ldots, A_r$ in a single trial.

Limit theorem for multinomial probabilities

Let

$$x_j = \frac{k_j - np_j}{\sqrt{np_j(1-p_j)}}$$

$$\mathrm{multi}(k_1, \ldots, k_r; n, p_1, \ldots, p_r) \sim$$

$$\sim \left((2\pi n)^{\frac{r-1}{2}} \sqrt{p_1 p_2 \cdots p_r}\right)^{-1} e^{-\frac{1}{2}\sum\limits_{j=1}^{r}(1-p_j)x_j^2}$$

Let a, b be fixed numbers. Put $\Delta = (\sqrt{np(1-p)})^{-1}$. If we substitute $\Delta \cdot \varphi(x_k)$ into a sum $\sum_{k=a}^{b} b(k; n, p)$, we obtain an integral sum

$$\sum_{k=a}^{b} \Delta \cdot \varphi(x_k)$$

for the definite integral

$$\int_{c_1}^{c_2} \varphi(u)\, du, \quad c_1 = \frac{a - np}{\sqrt{np(1-p)}}, \quad c_2 = \frac{b - np}{\sqrt{np(1-p)}}.$$

Integral theorem of De Moivre – Laplace

As $n \to \infty$

$$\sum_{k=a}^{b} b(k;n,p) \to \int_{c_1}^{c_2} \varphi(u)\,du$$

where $c_1 = \frac{a-np}{\sqrt{np(1-p)}}$, $c_2 = \frac{b-np}{\sqrt{np(1-p)}}$.

Define

$$F_n(x) = \sum_{k \leqslant np + x\sqrt{np(1-p)}} b(k;n,p), \quad \Phi(x) = \int_{-\infty}^{x} \varphi(u)\,du.$$

Special case of Berry-Esseen theorem

$$\sup_{-\infty \leqslant x \leqslant \infty} |F_n(x) - \Phi(x)| \leqslant \frac{p^2 + (1-p)^2}{\sqrt{np(1-p)}}$$

Useful properties of $\Phi(x)$:

$$\lim_{x\to-\infty}\Phi(x)=0,\qquad \lim_{x\to\infty}\Phi(x)=1,$$

$$\Phi(x)=\int\limits_{-\infty}^{x}\varphi(u)\,du=1-\Phi(-x),$$

$$\Phi_0(x)=\int\limits_{0}^{x}\varphi(u)\,du,\quad \Phi_0(x)=-\Phi_0(-x),$$

$$\Phi(0)=0{,}5,\quad \Phi(x)=0{,}5+\Phi_0(x).$$

Both $\Phi(x)$ and $\Phi_0(x)$ are tabulated.

Problem. During a data transmission each bit can be errorneously transmitted with probability 0,005. 10 000 bits are transmitted. What is the probability that at most 70 errors have occurred.

Problem. During a data transmission each bit can be errorneously transmitted with probability 0,005. 10 000 bits are transmitted. What is the probability that at most 70 errors have occurred.
Solution. Denote by X the number of errors during transmission. X has binomial distribution. Here $n = 10\,000$, $p = 0{,}005$. Use De Moivre – Laplace theorem:

$$np = 50, \quad np(1-p) = 49{,}75,$$

$$\mathrm{P}(X \leqslant 70) = \sum_{k=0}^{70} b(k;n,p) \approx$$

$$\approx \Phi((70-50)/\sqrt{49{,}75}) - \Phi((0-50)/\sqrt{49{,}75}) \approx$$

$$\approx \Phi_0(2{,}84) + \Phi_0(7{,}09) = 0{,}497744 + 0{,}5 = 0{,}997744.$$

Exact value: $\mathrm{P}(X \leqslant 70) = 0{,}9970970$.

Problem. What's the probability that in $n = 400$ tosses of a symmetric coin Heads appears between $m_1 = 180$ and $m_2 = 210$ times.

Problem. What's the probability that in $n = 400$ tosses of a symmetric coin Heads appears between $m_1 = 180$ and $m_2 = 210$ times.
Solution. Denote by X the number of Heads. X has binomial distribution with $n = 400$, $p = 0{,}5$.

$$np = 200, \quad \sqrt{np(1-p)} = 10,$$

$$\mathrm{P}(180 \leqslant X \leqslant 210) = \sum_{k=180}^{210} b(k; n, p) \approx$$

$$\approx \Phi((210 - 200)/10) - \Phi((180 - 200)/10) =$$

$$= \Phi_0(1) + \Phi_0(2) = 0{,}341345 + 0{,}477225 = 0{,}818570.$$

Exact value: $\mathrm{P}(180 \leqslant X \leqslant 210) = 0{,}8330306$.

In Example on screen 2 we had to find the probability

$$P = \sum_{|k-2{,}7\times 10^{23}|\geqslant 2{,}7\times 10^{13}} b(k; 5{,}4 \times 10^{23}, 1/2).$$

By virtue of De Moivre – Laplace theorem,

$$P \approx 2 \int\limits_{\frac{2{,}7\times 10^{13}}{\sqrt{5.4\times 10^{23}\times\frac{1}{4}}}}^{\infty} \varphi(u)\, du$$

Since

$$\int\limits_{z}^{\infty} e^{-\frac{u^2}{2}}\, du < \frac{1}{z}\int\limits_{z}^{\infty} ue^{-\frac{u^2}{2}}\, du = \frac{1}{z}e^{-\frac{z^2}{2}}$$

it follows that

$$P < \frac{1}{\sqrt{2\pi}\times 10}e^{2{,}7\times 100} < 10^{-100}.$$

Problem. A movie theatre is planned to have two cloackrooms of equal capacity. Assuming there are 1000 seats in the theatre and every person chooses between the cloackrooms independently of the others and equiprobably, what capacity should it be so that one of them overflows in no more than 5 % of cases?

Problem. A movie theatre is planned to have two cloackrooms of equal capacity. Assuming there are 1000 seats in the theatre and every person chooses between the cloackrooms independently of the others and equiprobably, what capacity should it be so that one of them overflows in no more than 5 % of cases?

Solution. Denote by N the capacity of one cloackroom. Lex X denote the number of people choosing the first cloackroom. Then there's no overflow when

$$X \leqslant N, \quad 1000 - X \leqslant N$$

whence

$$1000 - N \leqslant X \leqslant N.$$

We have Bernoulli trials with $n = 1000$ and $p = 1/2$. By virtue of De Moivre—Laplace theorem the probability of this event is

$$\Phi((N - np)/\sqrt{np(1-p)}) - \Phi((1000 - N - np)/\sqrt{np(1-p)}) =$$
$$= \Phi((N - 500)/5\sqrt{10}) - \Phi((500 - N)/5\sqrt{10})$$

Obviously, $N \geqslant 500$, so $N - 500 \geqslant 0$ and $500 - N \leqslant 0$. Thus, the desired probability is

$$2\Phi_0((N - 500)/5\sqrt{10}) \geqslant 0{,}95.$$

From the table,

$$\frac{N - 500}{5\sqrt{10}} \geqslant 1{,}96$$

and $N > 530{,}99$, so $N = 531$.