Áèíîìèàëüíîå ðàñïðåäåëåíèå

Îáîçíà÷åíèå	$\beta(m \mid n, p)$
Îáëàñòü çíà÷åíèé	$0 \le m \le n$, ãäå m – öåëîå
Ïàðàìåòðû	n – öåëîå ïîëîæèòåëüíîå ÷èñëî (èñïûòàíèé), $p \in (0,1)$ – ïàðàìåòð ñõåìû Áåðíóëëè (âåðîÿòíîñòü "óñïåõà"). Ïîìíèòå? âåëè÷èíó 1-p ïðèíÿòî îáîçíà÷àòü áóêâîé q.
Ïëîòíîñòü (ôóíêöèÿ âåðîÿòíîñòè)	Ïëîòíîñòü äèñêðåòíà: $\binom{n}{m} p^m q^{n-m}$. Çäåñü $\binom{n}{m}$ – ÷èñëî ñî÷åòàíèé èç n ýëåìåíòîâ ïî m, ïðè÷åì $m \in [0, n]$.
Ìàòåìàòè÷åñêîå îæèäàíèå	np
Äèñïåðñèÿ	npq
Ôóíêöèÿ ðàñïðåäåëåíèÿ	$\beta(m \mid n, p) = \sum_{x=0}^{m} \binom{n}{x} p^x q^{n-x}$

Ïîëåçíûå ñâîéñòâà

1. Êàê èçâåñòíî, ôóíêöèè áèíîìèàëüíîãî è áåòà- ðàñïðåäåëåíèé ñâÿçàíû ñëåäóþùèì ñîîòíîøåíèåì: $\beta(m \mid n, p) = B(1 - p \mid n - m, m + 1)$.
2. Ñèììåòðè÷íîñòè áåòà-ðàñïðåäåëåíèÿ ñîîòâåòñòâóåò ñèììåòðè÷íîñòü õâîñòîâ ðàñïðåäåëåíèÿ áèíîìèàëüíîãî: $\beta(m \mid n, p) = 1 - \beta(n - m - 1 \mid n, 1 - p)$.
3. Ñóììà k íåçàâèñèìûõ ñëó÷àéíûõ âåëè÷èí $\beta(m \mid n_i, p)$ åñòü òàêæå áèíîìèàëüíàÿ ñëó÷àéíàÿ âåëè÷èíà $\beta(m \mid n, p)$, ó êîòîðîé $n = \sum_{i=1}^{k} n_i$.
4. Ñîãëàñíî òåîðåìå Ìóàâðà-Ëàïëàñà ïðè $n \to \infty$ áèíîìèàëüíîå ðàñïðåäåëåíèå ñòðåìèòñÿ ê íîðìàëüíîìó. Âîò ñòàíäàðòíàÿ ôîðìóëèðîâêà:

   åñëè $npq > 5$ è $0.1 \le p \le 0.9$, òî $\beta(m|n,p) \approx \Phi\left(\frac{m - np + 0.5}{\sqrt{npq}}\right)$, ãäå Φ – ñòàíäàðòíîå íîðìàëüíîå ðàñïðåäåëåíèå.

   Âî âñåõ ó÷åáíèêàõ ãîâîðèòñÿ, ÷òî åñëè $npq > 25$, òî ýòó àïïðîêñèìàöèþ ìîæíî ïðèìåíÿòü ïðè ïðîèçâîëüíûõ çíà÷åíèÿõ p. Åñëè æå çíà÷åíèå p ìàëî́, òî áèíîìèàëüíîå ðàñïðåäåëåíèå ïðèíÿòî àïïðîêñèìèðîâàòü ïóàññîíîâñêèì: $\beta(m|n,p) \approx P(m \mid np)$. Ñ÷èòàåòñÿ, ÷òî ýòó ïîñëåäíþþ àïïðîêñèìàöèþ ñëåäóåò ïðèìåíÿòü ïðè $p < 0.1$.

Ãåíåðàöèÿ ñëó÷àéíûõ ÷èñåë

Áåðåì n íåçàâèñèìûõ ñëó÷àéíûõ ÷èñåë, ðàñïðåäåëåííûõ ðàâíîìåðíî íà îòðåçêå [0,1]. Êîëè÷åñòâî òåõ èç íèõ, êîòîðûå ìåíüøå p, çàäàþò ñëó÷àéíîå ÷èñëî, ïîä÷èíÿþùååñÿ ðàñïðåäåëåíèþ $\beta(m \mid n,p)$.

Âû÷èñëåíèå ôóíêöèè ðàñïðåäåëåíèÿ è åå êâàíòèëåé

Êîíå÷íî, ïðè âû÷èñëåíèè êóìóëÿòèâíîé ôóíêöèè ðàñïðåäåëåíèÿ ñëåäóåò âîñïîëüçîâàòüñÿ óïîìÿíóòîé ñâÿçüþ áèíîìèàëüíîãî è áåòà-ðàñïðåäåëåíèÿ. Ýòîò ñïîñîá çàâåäîìî ëó÷øå íåïîñðåäñòâåííîãî ñóììèðîâàíèÿ, êîãäà $n > 10$.

Â êëàññè÷åñêèõ ó÷åáíèêàõ ïî ñòàòèñòèêå äëÿ ïîëó÷åíèÿ çíà÷åíèé áèíîìèàëüíîãî ðàñïðåäåëåíèÿ ÷àñòî ðåêîìåíäóþò èñïîëüçîâàòü ôîðìóëû, îñíîâàííûå íà ïðåäåëüíûõ òåîðåìàõ (òèïà ôîðìóëû Ìóàâðà-Ëàïëàñà). Íåîáõîäèìî îòìåòèòü, ÷òî ñ ÷èñòî âû÷èñëèòåëüíîé òî÷êè çðåíèÿ öåííîñòü ýòèõ òåîðåì áëèçêà ê íóëþ, îñîáåííî ñåé÷àñ, êîãäà ïðàêòè÷åñêè íà êàæäîì ñòîëå ñòîèò ìîùíûé êîìïüþòåð. Îñíîâíîé íåäîñòàòîê ïðèâåäåííûõ àïïðîêñèìàöèé – èõ ñîâåðøåííî íåäîñòàòî÷íàÿ òî÷íîñòü ïðè çíà÷åíèÿõ n, õàðàêòåðíûõ äëÿ áîëüøèíñòâà ïðèëîæåíèé. Íå ìåíüøèì íåäîñòàòêîì ÿâëÿåòñÿ è îòñóòñòâèå ñêîëüêî-íèáóäü ÷åòêèõ ðåêîìåíäàöèé î ïðèìåíèìîñòè òîé èëè èíîé àïïðîêñèìàöèè (â ñòàíäàðòíûõ òåêñòàõ ïðèâîäÿòñÿ ëèøü àñèìïòîòè÷åñêèå ôîðìóëèðîâêè, îíè íå ñîïðîâîæäàþòñÿ îöåíêàìè òî÷íîñòè è, ñëåäîâàòåëüíî, ìàëî ïîëåçíû). ß áû ñêàçàë, ÷òî îáå ôîðìóëû ïðèãîäíû ëèøü ïðè $n < 200$ è äëÿ ñîâñåì ãðóáûõ, îðèåíòèðîâî÷íûõ ðàñ÷åòîâ, ïðè÷åì äåëàåìûõ “âðó÷íóþ” ñ ïîìîùüþ ñòàòèñòè÷åñêèõ òàáëèö. À âîò ñâÿçü ìåæäó áèíîìèàëüíûì ðàñïðåäåëåíèåì è áåòà-ðàñïðåäåëåíèåì ïîçâîëÿåò âû÷èñëÿòü áèíîìèàëüíîå ðàñïðåäåëåíèå äîñòàòî÷íî ýêîíîìíî.

ß íå ðàññìàòðèâàþ çäåñü çàäà÷ó ïîèñêà êâàíòèëåé: äëÿ äèñêðåòíûõ ðàñïðåäåëåíèé îíà òðèâèàëüíà, à â òåõ çàäà÷àõ, ãäå òàêèå ðàñïðåäåëåíèÿ âîçíèêàþò, îíà, êàê ïðàâèëî, è íå àêòóàëüíà. Åñëè æå êâàíòèëè âñå-òàêè ïîíàäîáÿòñÿ, ðåêîìåíäóþ òàê ïåðåôîðìóëèðîâàòü çàäà÷ó, ÷òîáû ðàáîòàòü ñ p-çíà÷åíèÿìè (íàáëþäàåìûìè çíà÷èìîñòÿìè). Âîò ïðèìåð: ïðè ðåàëèçàöèè íåêîòîðûõ ïåðåáîðíûõ àëãîðèòìîâ íà êàæäîì øàãå òðåáóåòñÿ ïðîâåðÿòü ñòàòèñòè÷åñêóþ ãèïîòåçó î áèíîìèàëüíîé ñëó÷àéíîé âåëè÷èíå. Ñîãëàñíî êëàññè÷åñêîìó ïîäõîäó íà êàæäîì øàãå íóæíî âû÷èñëèòü ñòàòèñòèêó êðèòåðèÿ è ñðàâíèòü åå çíà÷åíèå ñ ãðàíèöåé êðèòè÷åñêîãî ìíîæåñòâà. Ïîñêîëüêó, îäíàêî, àëãîðèòì ïåðåáîðíûé, ïðèõîäèòñÿ îïðåäåëÿòü ãðàíèöó êðèòè÷åñêîãî ìíîæåñòâà êàæäûé ðàç çàíîâî (âåäü îò øàãà ê øàãó îáúåìû âûáîðêè ìåíÿþòñÿ), ÷òî íåïðîèçâîäèòåëüíî óâåëè÷èâàåò âðåìåííûå çàòðàòû. Ñîâðåìåííûé ïîäõîä ðåêîìåíäóåò âû÷èñëÿòü íàáëþäàåìóþ çíà÷èìîñòü è ñðàâíèâàòü åå ñ äîâåðèòåëüíîé âåðîÿòíîñòüþ, ýêîíîìÿ íà ïîèñêå êâàíòèëåé.

Ïîýòîìó â ïðèâîäèìûõ íèæå êîäàõ îòñóòñòâóåò âû÷èñëåíèå îáðàòíîé ôóíêöèè, âçàìåí ïðèâåäåíà ôóíêöèÿ `rev_binom`, êîòîðàÿ ïî çàäàííîìó êîëè÷åñòâó n èñïûòàíèé, ÷èñëó m óñïåõîâ â íèõ è çíà÷åíèþ y âåðîÿòíîñòè ïîëó÷èòü ýòè m óñïåõîâ âû÷èñëÿåò âåðîÿòíîñòü p óñïåõà â îòäåëüíîì èñïûòàíèè. Ïðè ýòîì èñïîëüçóåòñÿ âûøåóïîìÿíóòàÿ ñâÿçü ìåæäó áèíîìèàëüíûì è áåòà ðàñïðåäåëåíèÿìè.

Ôàêòè÷åñêè, ýòà ôóíêöèÿ ïîçâîëÿåò ïîëó÷àòü ãðàíèöû äîâåðèòåëüíûõ èíòåðâàëîâ. Â ñàìîì äåëå, ïðåäïîëîæèì, ÷òî â n áèíîìèàëüíûõ èñïûòàíèÿõ ìû ïîëó÷èëè m óñïåõîâ. Êàê èçâåñòíî, ëåâàÿ ãðàíèöà äâóõñòîðîííåãî äîâåðèòåëüíîãî èíòåðâàëà äëÿ ïàðàìåòðà p ñ äîâåðèòåëüíûì óðîâíåì α ðàâíà 0, åñëè $m = 0$, à äëÿ $0 < m \le n$ ÿâëÿåòñÿ ðåøåíèåì óðàâíåíèÿ $\Pr\{\mu \ge m \mid \pi\} = \frac{\alpha}{2}$. Àíàëîãè÷íî, ïðàâàÿ ãðàíèöà ðàâíà 1, åñëè $m = n$, à äëÿ $0 \le m < n$ ÿâëÿåòñÿ ðåøåíèåì óðàâíåíèÿ $\Pr\{\mu \le m \mid \pi\} = \frac{\alpha}{2}$. Îòñþäà âûòåêàåò, ÷òî äëÿ ïîèñêà ëåâîé ãðàíèöû ìû äîëæíû ðåøàòü îòíîñèòåëüíî π óðàâíåíèå

$$B(\pi \mid m, n - m + 1) = \frac{1 - \alpha}{2},$$

à äëÿ ïîèñêà ïðàâîé – óðàâíåíèå

$$B(\pi \mid m + 1, n - m) = \frac{1 + \alpha}{2}.$$

Îíè è ðåøàþòñÿ â ôóíêöèÿõ `leftCI_binom` è `rightCI_binom`, âîçâðàùàþùèõ âåðõíþþ è íèæíþþ ãðàíèöû äâóõñòîðîííåãî äîâåðèòåëüíîãî èíòåðâàëà ñîîòâåòñòâåííî.

Õî÷ó çàìåòèòü, ÷òî åñëè íå íóæíà ñîâñåì óæ íåèìîâåðíàÿ òî÷íîñòü, òî ïðè äîñòàòî÷íî áîëüøèõ n ìîæíî âîñïîëüçîâàòüñÿ ñëåäóþùåé àïïðîêñèìàöèåé [Á.Ë. âàí äåð Âàðäåí, Ìàòåìàòè÷åñêàÿ ñòàòèñòèêà. Ì.: ÈË, 1960, ãë. 2, ðàçä. 7]:

$$\frac{m + \frac{g^2}{2} \pm g\sqrt{\frac{m(n - m)}{n} + \frac{g^2}{4}}}{n + g^2},$$

ãäå g – êâàíòèëü íîðìàëüíîãî ðàñïðåäåëåíèÿ. Öåííîñòü ýòîé àïïðîêñèìàöèè â òîì, ÷òî èìåþòñÿ î÷åíü ïðîñòûå ïðèáëèæåíèÿ, ïîçâîëÿþùèå âû÷èñëÿòü êâàíòèëè íîðìàëüíîãî ðàñïðåäåëåíèÿ (ñì. òåêñò î âû÷èñëåíèè íîðìàëüíîãî ðàñïðåäåëåíèÿ è ñîîòâåòñòâóþùèé ðàçäåë äàííîãî ñïðàâî÷íèêà). Â ìîåé ïðàêòèêå (â îñíîâíîì, ïðè $n > 100$) ýòà àïïðîêñèìàöèÿ äàâàëà ïðèìåðíî 3-4 çíàêà, ÷åãî, êàê ïðàâèëî, âïîëíå äîñòàòî÷íî.

Äëÿ âû÷èñëåíèé ñ ïîìîùüþ íèæåñëåäóþùèõ êîäîâ ïîòðåáóþòñÿ ôàéëû `betaDF.h`, `betaDF.cpp` (ñì. ðàçäåë î áåòà-ðàñïðåäåëåíèè), à òàêæå `logGamma.h`, `logGamma.cpp` (ñì. ïðèëîæåíèå À).

Ôàéë BinomialDF.h

```
#ifndef __BINOMIAL_H__

#include "betaDF.h"

double binomialDF(double n, double m, double p);
/*
 * Ïóñòü èìååòñÿ 'n' íåçàâèñèìûõ íàáëþäåíèé
 *      ñ âåðîÿòíîñòüþ 'p' óñïåõà â êàæäîì.
 * Âû÷èñëÿåòñÿ âåðîÿòíîñòü b(m|n,p) òîãî, ÷òî ÷èñëî óñïåõîâ çàêëþ÷åíî
 * ìåæäó 0 è 'm' (âêëþ÷èòåëüíî).
 */

double rev_binom(double n, double m, double y);
/*
```

```
 * Ïóñòü èçâåñòíà âåðîÿòíîñòü y íàñòóïëåíèÿ íå ìåíåå m óñïåõîâ
 * â n èñïûòàíèÿõ ñõåìû Áåðíóëëè. Ôóíêöèÿ íàõîäèò âåðîÿòíîñòüp
 * óñïåõà â îòäåëüíîì èñïûòàíèè.
 */

double leftCI_binom(double n, double m, double level);
/* Ïóñòü èìååòñÿ 'n' íåçàâèñèìûõ íàáëþäåíèé
 *      ñ âåðîÿòíîñòüþ 'p' óñïåõà â êàæäîì.
 * Âû÷èñëÿåòñÿ ëåâàÿ ãðàíèöà äîâåðèòåëüíîãî èíòåðâàëà
 * ñ óðîâíåì çíà÷èìîñòè y.
 */
double rightCI_binom(double n, double m, double level);
/* Ïóñòü èìååòñÿ 'n' íåçàâèñèìûõ íàáëþäåíèé
 *      ñ âåðîÿòíîñòüþ 'p' óñïåõà â êàæäîì.
 * Âû÷èñëÿåòñÿ ïðàâàÿ ãðàíèöà äîâåðèòåëüíîãî èíòåðâàëà
 * ñ óðîâíåì çíà÷èìîñòè y.
 */

#endif              /* Ends #ifndef __BINOMIAL_H__  */
```

Ôàéë BinomialDF.cpp

```
#include <math.h>
#include <assert.h>

#include "betaDF.h"

ENTRY double
binomialDF(double n, double m, double p)
/*
 * Ïóñòü èìååòñÿ 'n' íåçàâèñèìûõ íàáëþäåíèé
 *      ñ âåðîÿòíîñòüþ 'p' óñïåõà â êàæäîì.
 * Âû÷èñëÿåòñÿ âåðîÿòíîñòü b(m|n,p) òîãî, ÷òî ÷èñëî óñïåõîâ çàêëþ÷åíî
 * ìåæäó 0 è 'm' (âêëþ÷èòåëüíî), ò.å.
 * ñóììó áèíîìèàëüíûõ âåðîÿòíîñòåé îò 0 äî m:
 *
 *   m
 *   --  ( n )   j      n-j
 *   >   (   )  p  (1-p)
 *   --  ( j )
 *  j=0
 *
 * Âû÷èñëåíèÿ íå ïîäðàçóìåâàþò òóïîå ñóììèðîâàíèå - èñïîëüçóåòñÿ
 * ñëåäóþùàÿ ñâÿçü áèíîìèàëüíîãî ðàñïðåäåëåíèÿ ñ
 * öåíòðàëüíûì áåòà-ðàñïðåäåëåíèåì:
 *
 *    b(m|n,p) = Beta(1-p|n-m,m+1).
 *
 * Àðãóìåíòû äîëæíû áûòü ïîëîæèòåëüíûìè, ïðè÷åì 0 <=p <= 1.
 */
{
   assert((n > 0) && (p >= 0) && (p <= 1));
   if (m < 0)
      return 0;
   else if (m == 0)
      return pow(1-p, n);
   else if (m >= n)
      return 1;
   else
```

```
      return BetaDF(n-m, m+1).value(1-p);
}/* binomialDF */

ENTRY double
rev_binom(double n, double m, double y)
/*
 * Ïóñòü èçâåñòíà âåðîÿòíîñòü y íàñòóïëåíèÿ íå ìåíåå 'm’ óñïåõîâ
 * â 'n' èñïûòàíèÿõ ñõåìû Áåðíóëëè. Ôóíêöèÿ íàõîäèò âåðîÿòíîñòü 'p’
 * óñïåõà â îòäåëüíîì èñïûòàíèè.
 *
 * Â âû÷èñëåíèÿõ èñïîëüçóåòñÿ ñëåäóþùåå ñîîòíîøåíèå
 *
 *        1 - p = rev_Beta(y|n-m,m+1).
 */
{
   assert( (n > 0) && (m >= 0) && (m <= n) && (y >=0) && (y <= 1) );
   return 1-BetaDF(n-m, m+1).inv(y);
}/*rev_binom*/

ENTRY double
leftCI_binom(double n, double m, double y)
/* Ïóñòü èìååòñÿ 'n' íåçàâèñèìûõ íàáëþäåíèé
 *      ñ âåðîÿòíîñòüþ 'p' óñïåõà â êàæäîì.
 * Âû÷èñëÿåòñÿ ëåâàÿ ãðàíèöà äâóõñòîðîííåãî äîâåðèòåëüíîãî èíòåðâàëà
 * ñ óðîâíåì çíà÷èìîñòè 'y', êîòîðàÿ, ñîãëàñíî îïðåäåëåíèþ, ÿâëÿåòñÿ
 * ðåøåíèåì óðàâíåíèÿ
 *
 */
{
   assert( (n > 0) && (m >= 0) && (m <= n) && (y >= 0.5) && (y < 1) );
   return BetaDF(m, n-m+1).inv(0.5*(1-y));
}/*leftCI_binom*/

ENTRY double
rightCI_binom(double n, double m, double y)
/* Ïóñòü èìååòñÿ 'n' íåçàâèñèìûõ íàáëþäåíèé
 *      ñ âåðîÿòíîñòüþ 'p' óñïåõà â êàæäîì.
 * Âû÷èñëÿåòñÿ ïðàâàÿ ãðàíèöà äîâåðèòåëüíîãî èíòåðâàëà
 * ñ óðîâíåì çíà÷èìîñòè y.
 */
{
   assert( (n > 0) && (m >= 0) && (m <= n) && (y >= 0.5) && (y < 1) );
   return BetaDF(m+1, n-m).inv((1+y)/2);
}/*rightCI_binom*/
```