Моделирование вырожденного гипергеометрического и бета распределений

Гуляева Екатерина Игоревна, 522-я группа

Санкт-Петербургский государственный университет
Математико-механический факультет
Кафедра статистического моделирования

Научный руководитель — к. ф.-м. н. **доцент В. Б. Христинич**
Рецензент — д. ф.-м. н. **профессор С. М. Ермаков**

Санкт-Петербург
2014г.

Основные цели дипломной работы. Постановка задач

Основные цели дипломной работы:

- построение эффективных методов моделирования $Beta$ и вырожденного гипергеометрического (CH) распределений;
- использование последних предложенных в литературе модификаций методов моделирования;
- изучение и вычисление основных характеристик исследуемых распределений.

Постановка задач:

- изучение специальных функций, необходимых для анализа $Beta$ и CH распределений;
- построение связей между специальными функциями и изучаемыми распределениями, включая их плотности, функции распределений, моменты и характеристические функции;
- изучение моделируемых распределений;
- исследование связи $Beta$ и CH распределений;
- разработка алгоритмов и программ для вычисления значений специальных функций с учетом их особенностей, применительно к рассматриваемым распределениям.

Бета-функция:

$$B(p,q)=\int\limits_0^1 t^{p-1}(1-t)^{q-1}dt,\quad Re\ p>0,\quad Re\ q>0; \tag{1}$$

Неполная бета-функция:

$$B_x(p,q)=\int\limits_0^x t^{p-1}(1-t)^{q-1}dt,\quad Re\ p>0,\quad Re\ q>0,\quad 0\le x\le 1; \tag{2}$$

Гипергеометрическая функция:

$$\begin{gathered}{}_2F_1(a,b;c;x)=\sum_{n=0}^{\infty}\frac{(a)_n(b)_n x^n}{(c)_n n!},\quad c\ne 0,-1,-2,...;\\ (a)_0=1,(a)_n=a(a+1)...(a+n-1);\end{gathered} \tag{3}$$

$${}_2F_1(a,b;c;x)=\frac{1}{B(c,c-b)}\int\limits_0^1 u^{a-1}(1-u)^{c-a-1}(1-xu)^{-b}du, Re\ c>Re\ b>0,\quad |\arg(1-x)|<\pi;$$

Вырожденная гипергеометрическая функция:

$$\begin{gathered}{}_1F_1(a;c;x)=\sum_{n=0}^{\infty}\frac{(a)_n x^n}{(c)_n n!},\quad c\ne 0,-1,-2,...;\\ {}_1F_1(a;c;x)=\frac{1}{B(c,c-a)}\int\limits_0^1 u^{a-1}(1-u)^{c-a-1}e^{xu}du,\quad Re\ c>Re\ a>0.\end{gathered} \tag{4}$$

Условия сходимости ${}_2F_1(a,b;c;x)$:

$$\begin{cases} |x| < 1; \\ |x| = 1, \quad Re(c-a-b) > 0; \\ |x| = 1, \quad x \neq 1, \quad -1 < Re(c-a-b) \leq 0. \end{cases}$$

Переход от аргумента x к $(1-x)$:

$$\begin{aligned} {}_2F_1(a,b;c;x) &= \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}\,{}_2F_1(a,b;1+a+b-c;1-x) + \\ &+ \frac{\Gamma(c)\Gamma(a+b-c)}{\Gamma(a)\Gamma(b)}(1-x)^{c-a-b}\,{}_2F_1(c-a,c-b;1-a-b+c;1-x), \\ &\qquad c-a-b \neq 0, \pm 1, \pm 2\ldots, \quad |arg(1-x)| < \pi. \end{aligned} \tag{5}$$

Если $a=-n$ либо $b=-n$, то (5) приобретает вид:

$${}_2F_1(a,b;c;x) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}\,{}_2F_1(a,b;1+a+b-c;1-x),$$

а в случаях $c-a=-n$ либо $c-b=-n$:

$${}_2F_1(a,b;c;x) = \frac{\Gamma(c)\Gamma(a+b-c)}{\Gamma(a)\Gamma(b)}(1-x)^{c-a-b}\,{}_2F_1(c-a,c-b;1-a-b+c;1-x).$$

Условия сходимости ${}_1F_1(a;c;x)$: $|x| < \infty$.
Переход от аргумента x к $(-x)$:

$$ {}_1F_1(a;c;x) = \exp(x)\,{}_1F_1(c-a;c;-x). \tag{6} $$

Частные случаи:

$$ {}_1F_1(-n;m;x) = \sum_{k=0}^{n} \frac{(-n)_k x^k}{(m)_k k!}, \quad n \in \mathbb{N}, m \in \mathbb{N}^+; $$
$$ {}_1F_1(-n;-m;x) = \sum_{k=0}^{n} \frac{(-n)_k x^k}{(-m)_k k!}, \quad n \in \mathbb{N}, m \in \mathbb{Z}^+, m \geq n. \tag{7} $$

При $x \to \infty$ необходимо пользоваться разложением Пуанкаре:

$$ {}_1F_1(a;c;x) = \frac{e^x x^{(a-c)}}{\Gamma(a)} \sum_{k=0}^{\infty} \frac{(1-a)_k (b-a)_k}{k!} x^{(-k)}. \tag{8} $$

Связи между специальными функциями, рассмотренными в работе

Неполная гамма-функция: $\gamma(a,x) = \int\limits_0^x e^{-t} t^{a-1} dt;$

Дигамма-функция: $\psi(x) = \dfrac{d \ln \Gamma(x)}{dx}.$

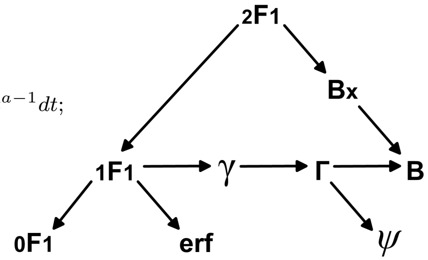


Рис.1: Связи между специальными функциями.

$$B_x(p,q) = \frac{x^p}{p} {}_2F_1(p, 1-q; p+1; x);$$

$$\gamma(a,x) = \frac{x^a}{a} {}_1F_1(a; a+1; -x);$$

$$\psi(x) = \frac{d \ln \Gamma(x)}{dx};$$

$${}_1F_1(a;c;x) = \lim_{b\to\infty} {}_2F_1(a,b;c;\frac{x}{b});$$

$$erf(x) = x {}_1F_1(\frac{1}{2}; \frac{3}{2}; -x^2);$$

$$\lim_{a\to\infty} {}_1F_1(a;b;\frac{x}{a}) = \frac{2^{(}1-b)}{\Gamma(b)} {}_0F_1(\cdot; b; x).$$

Плотность бета-распределения:

$$Beta(\alpha,\beta,x)=\frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1},\quad 0<x<1; \tag{9}$$

Функция распределения:

$$I_x(\alpha,\beta)\equiv F(\alpha,\beta,x)=\frac{B_x(\alpha,\beta)}{B(\alpha,\beta)};$$

Математическое ожидание:

$$E(X)=\alpha/(\alpha+\beta);$$

Начальные моменты:

$$m_k(X)=\frac{\Gamma(\alpha+\beta)\Gamma(\alpha+k)}{\Gamma(\alpha)\Gamma(\alpha+\beta+k)};$$

Дисперсия бета-распределения:

$$D(X)=\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)};$$

Характеристическая функция:

$$\varphi_X(t)=\sum_{k=0}^{\infty}\frac{(\alpha)_k}{(\alpha+\beta)_k}\frac{(it)^k}{k!}={}_2F_1(\alpha;\alpha+\beta;it).$$

Метод отбор для моделирования $Beta$-распределения
Модификация Ченга(Cheng)

f - плотность, моделируемого распределения;
g - вспомогательная плотность;
G - функция распределения, соответствующая плотности g.

Алгоритм 1: метод отбора

1. Generate $U_1, U_2 \sim unif(0,1)$;
2. Set $A = \max\limits_{x>0} \frac{f}{g}$;
3. Set $X = G^{-1}(U_1)$;
4. If $\frac{f(X)}{Ag(X)} \leq U_2$ go to 1 else return X.

Алгоритм Метрополиса-Хастингса для моделирования случайных величин

f - плотность, моделируемого распределения;
q - вспомогательная плотность.

Алгоритм 2: метод Метрополиса-Хастингса

Given $x^{(t)}$,
1. Generate $Y_t \sim q(y|x^{(t)})$;
2. Set

$$x^{(t+1)} = \begin{cases} Y_t & \text{with probability} \quad \rho(x^{(t)}, Y_t), \\ x^{(t)} & \text{with probability} \quad 1 - \rho(x^{(t)}, Y_t); \end{cases}$$

where

$$\rho(x, y) = min\left\{\frac{f(y)}{f(x)} \frac{q(x|y)}{q(y|x)}, 1\right\}.$$

Сравнение методов моделирования $Beta$-распределения

Таблица 1: Сравнение метод для моделирования $Beta$-распределения

Алгоритм	Количество сл.вел.	Время (в миллисекундах)
Метод отбора	10 000	29
Модифицированный метод отбора	10 000	27
Метод Метрополиса-Хастингса	10 000	31
Метод отбора	1 000 000	1576
Модифицированный метод отбора	1 000 000	1050
Метод Метрополиса-Хастингса	1 000 000	1589

- В большинстве прикладных задач получаемая марковская цепь в алгоритме Метрополиса-Хастингса имеет стационарное распределение, совпадающее с распределением f. Таким образом, для получения независимой выборки необходимо, запустив цепочку, исключить некоторое количество начальных значений, чтобы цепочка успела сойтись к стационарному распределению;
- Метод Метрополиса-Хастингса включает в себя повторения в случае отказа от Y_t;
- Однако метод отбора включает в себя подсчет $max\frac{f(x)}{g(x)}$, что не требуется в алгоритме Метрополиса-Хастингса. Это является большим преимуществом, так как это вычисление может занимать довольно много времени или быть неточным.

CH распределение: определение и свойства

Плотность вырожденного гипергеометрического распределения:

$$f(\alpha,\beta,\gamma,x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{x^{\alpha-1}(1-x)^{\beta-1}e^{\gamma x}}{{}_1F_1(\alpha;\alpha+\beta;\gamma)}, \quad 0<x<1, \alpha>0, \beta>0, -\infty<\gamma<\infty \qquad (10)$$

Связь CH и $Beta$ распределений

$$f(\alpha,\beta,\gamma,x) = ({}_1F_1(\alpha;\alpha+\beta;\gamma))^{-1} \sum_{k=0}^{\infty} \frac{(\alpha)_k}{(\alpha+\beta)_k} \frac{(\gamma)^k}{k!} Beta(\alpha+k,\beta,x) \qquad (11)$$

Математическое ожидание:

$$E(X) = \frac{\alpha}{\alpha+\beta} \frac{{}_1F_1(\alpha+1;\alpha+\beta+1,\gamma)}{{}_1F_1(\alpha;\alpha+\beta,\gamma)};$$

Моменты вырожденного гипергеометрического распределения:

$$m_k(X) = \frac{(\alpha)_k}{(\alpha+\beta)_k} \frac{{}_1F_1(\alpha+k;\alpha+\beta+k,\gamma)}{{}_1F_1(\alpha;\alpha+\beta,\gamma)};$$

Дисперсия CH-распределения:

$$D(X) = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} \frac{{}_1F_1(\alpha+2;\alpha+\beta+2,\gamma)}{{}_1F_1(\alpha;\alpha+\beta,\gamma)} - (\frac{\alpha}{\alpha+\beta})^2 \left(\frac{{}_1F_1(\alpha+1;\alpha+\beta+1,\gamma)}{{}_1F_1(\alpha;\alpha+\beta,\gamma)}\right)^2;$$

Характеристическая функция:

$$\phi_X(t) = \frac{{}_1F_1(\alpha,\beta,it+\gamma)}{{}_1F_1(\alpha,\alpha+\beta,\gamma)}.$$

Функция распределения CH

$$F(\alpha,\beta,\gamma,x) = \left(B(\alpha,\beta)\,{}_1F_1(\alpha;\alpha+\beta;\gamma)\right)^{-1}\int\limits_0^x t^{\alpha-1}(1-t)^{\beta-1}e^{\gamma t}dt; \qquad (12)$$

$$F(\alpha,\beta,\gamma,x) = \left(B(\alpha,\beta)\,{}_1F_1(\alpha;\alpha+\beta,\gamma)\right)^{-1}\sum_{k=0}^{\infty}\frac{\gamma^k}{k!}B_x(\alpha+k,\beta). \qquad (13)$$

Разложение в окрестности 0:

$$F(\alpha,\beta,\gamma,x) = \left(B(\alpha,\beta)\,{}_1F_1(\alpha;\alpha+\beta,\gamma)\right)^{-1}x^{\alpha}(C_0+C_1x+C_2x^2+C_3x^3+O(x^4)), \qquad (14)$$

где x^{α} - выделенная особенность и

$$C_0 = \frac{1}{\alpha};$$

$$C_1 = \frac{1-\beta}{\alpha+1}+\frac{\gamma}{\alpha+1};$$

$$C_2 = \frac{(1-\beta)(2-\beta)}{2!(\alpha+2)}+\frac{\gamma(1-\beta)}{\alpha+2}+\frac{\gamma^2}{2!(\alpha+2)};$$

$$C_n = \sum_{k=0}^{n}\frac{(1-\beta)_{n-k}\gamma^k}{k!(n-k)!(\alpha+n)}.$$

Функция распределения CH

$$F(\alpha,\beta,\gamma,x) = \left(B(\alpha,\beta)\ {}_1F_1(\alpha;\alpha+\beta;\gamma)\right)^{-1} \int\limits_0^x t^{\alpha-1}(1-t)^{\beta-1} e^{\gamma t} dt;$$

$$F(\alpha,\beta,\gamma,x) = \left(B(\alpha,\beta)\ {}_1F_1(\alpha;\alpha+\beta,\gamma)\right)^{-1} \sum_{k=0}^{\infty} \frac{\gamma^k}{k!} B_x(\alpha+k,\beta).$$

Разложение в окрестности 1:

$$F(\alpha,\beta,\gamma,x) = \left(B(\alpha,\beta)\ {}_1F_1(\alpha;\alpha+\beta,\gamma)\right)^{-1}\left(P-(1-x)^{\beta}\,Q\right), \tag{15}$$

где $(1-x)^{\beta}$ - выделенная особенность и

$$P = \sum_{k=0}^{\infty} \frac{\gamma^k}{k!} B(\alpha+k,\beta);$$

$$Q = \sum_{l=0}^{\infty} C_l \frac{(1-x)^l}{(\beta+l)l!};$$

$$C_l = \sum_{m=0}^{\infty} \frac{\gamma^m}{m!} (1-\alpha-m)_l.$$

Алгоритм Метрополиса-Хастингса для моделирования CH распределения

f - плотность вырожденного гипергеометрического распределения;
$q = unif(0,1)$ - вспомогательная плотность.

Алгоритм 3: моделирование CH распределения

Algorithm:
Given $x^{(t)}$,
1. Generate $Y_t \sim unif(0,1)$;
2. Set

$$x^{(t+1)} = x^{(t)} + (Y_t - x^{(t)})(U(0,1) < \rho),$$

where

$$\rho = \frac{f(Y_t)}{f(x^{(t)})}.$$

Таблица 2: Метод Метрополиса-Хастингса для CH и $Beta$ распределений

Распределение	Количество сл.вел.	Время (в миллисекундах)
$Beta(2,3)$	10 000	31
$CH(1,2,3)$	10 000	193
$Beta(2,3)$	1 000 000	1589
$CH(1,2,3)$	1 000 000	16965

Результаты вычислений и моделирования

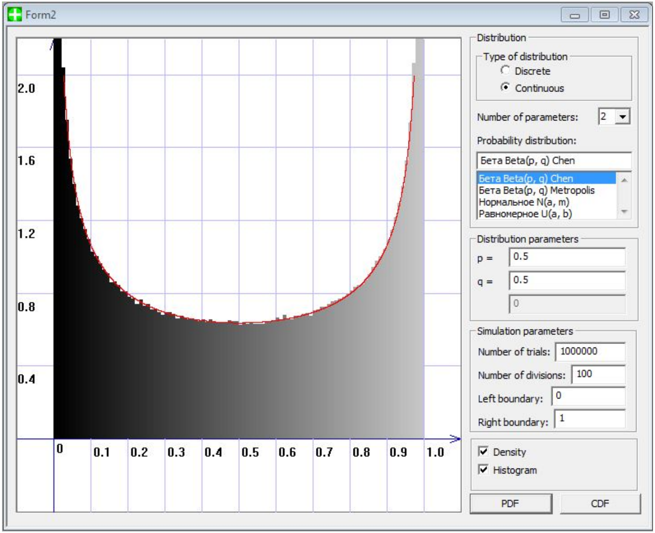


Рис.2: Плотность $Beta$ распределения с параметрами $\alpha = 0.5, \beta = 0.5$.

Результаты вычислений и моделирования

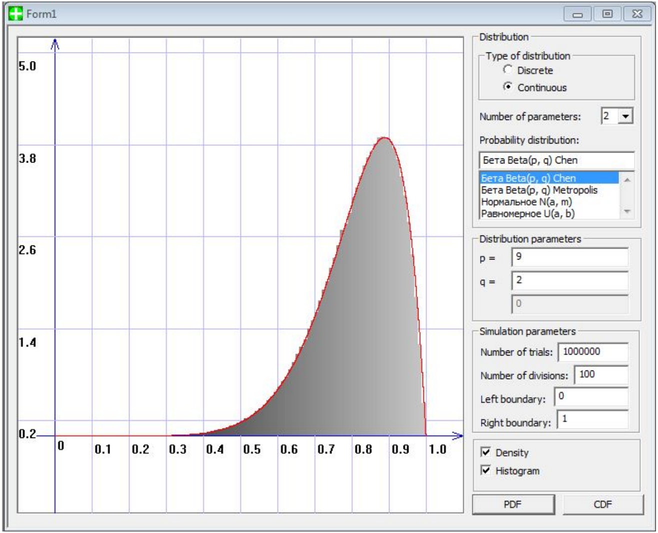


Рис.3: Плотность $Beta$ распределения с параметрами $\alpha = 9, \beta = 2$.

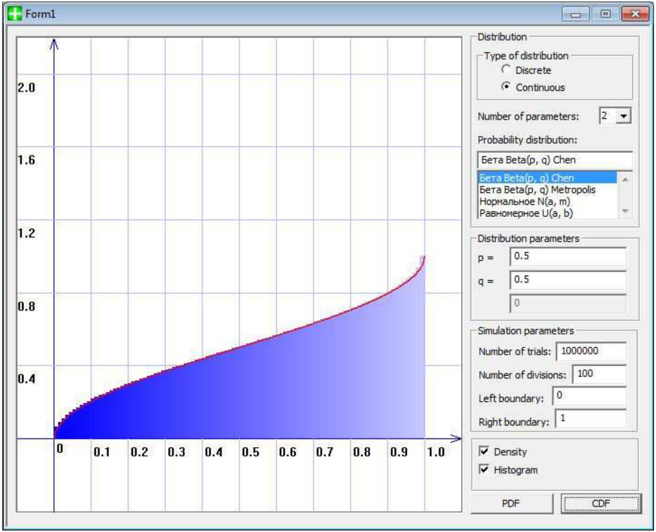


Рис.4: Функция $Beta$ распределения с параметрами $\alpha = 0.5, \beta = 0.5$.

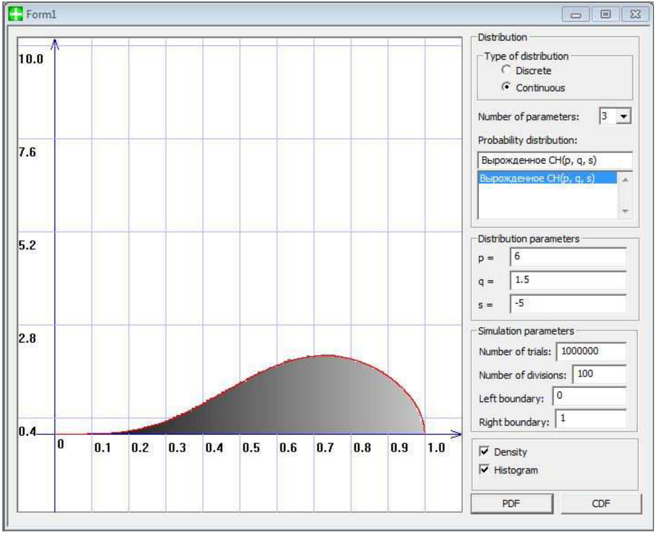


Рис.5: Плотность CH распределения с параметрами $\alpha = 6, \beta = 1.5, \gamma = -5$.

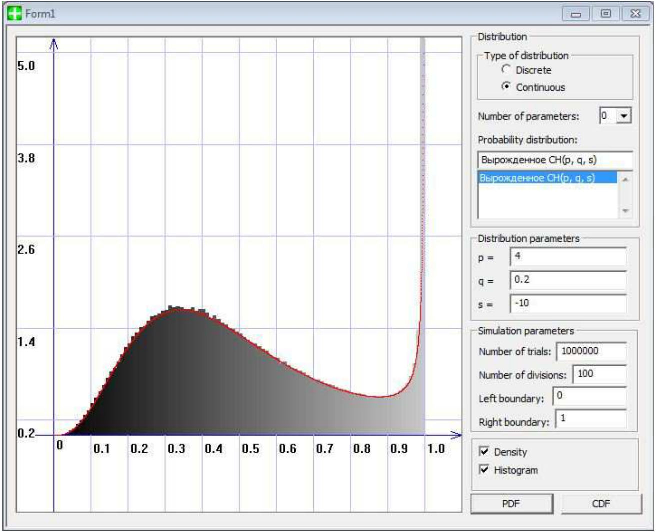


Рис.6: Плотность CH распределения с параметрами $\alpha = 4, \beta = 0.2, \gamma = -10$.

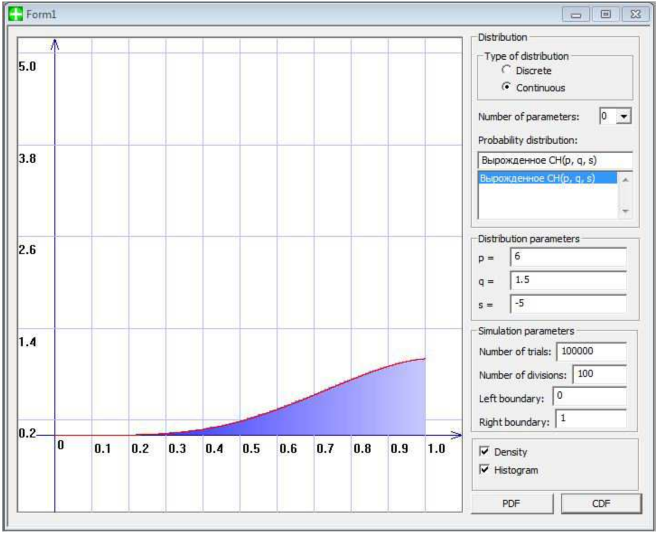
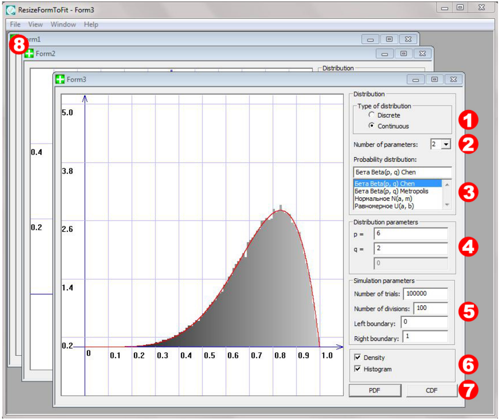


Рис.7: Функция CH распределения с параметрами $\alpha = 6, \beta = 1.5, \gamma = -5$.

Интерфейс программы



1) Тип распределения:
   (дискретное/непрерывное);
2) Кол-во параметров;
3) Распределение;
4) Параметры распределения;
5) Параметры моделирования:
   (границы, кол-во сл.вел,
   кол-во делений гистограммы)
6) Теоретическая кривая/гистограмма;
7) Плотность/функция распределения;
8) Глобальное меню программы.

Результаты работы

- Разработаны алгоритмы и программа, реализующие подсчет значений гипергеометрических функций, гамма и бета функций, неполной бета-функции, а также вычисление плотностей и функций распределения вероятностей бета и вырожденного гипергеометрического распределений;
- Моделирование бета-распределения осуществлено двумя рассмотренными в работе методами: методом отбора и методом Метрополиса-Хастингса;
- Вырожденное гипергеометрическое распределение моделируется универсальным методом Метрополиса-Хастингса;
- Для CH распределения представлены формула функции распределения вероятностей, а так же разложения ее в особых точках;
- В работе получены и приведены все основные характеристики для рассматриваемых распределений, такие как математическое ожидание, дисперсия, моменты, а также характеристическая функция.

Спасибо за внимание!