**Домашнее задание № 4.**

1. **Сравнительный анализ современных методов оптимизации (SGD, NAG, Adagrad, ADAM) на примере многослойного персептрона.**
2. **Использование генетического алгоритма для оптимизации гиперпараметров (число слоев и число нейронов) многослойного персептрона.**
3. **Сравнительный анализ современных методов оптимизации (SGD, NAG, Adagrad, ADAM) на примере многослойного персептрона.**
   1. **Stochastic gradient descent (SGD)**

**–**

В отличии от традиционного градиентного метода, метод стохастического градиента позволяет обрабатывать за одну итерацию относительно небольшой набор обучающих данных из всего набора данных (выборки - batch).

Применение этого метода позволяет избежать недостатки традиционного градиентного метода:

1. Требуется значительно меньше необходимого объема памяти.
2. Повышается скорость сходимости.
3. Увеличивается робастность к невыпуклости целевых функций.
4. Возможна оптимизация многоэкстремальных целевых функций.
   * 1. **Mini-batch gradient descent**

𝜃 = 𝜃 –𝜂 ∇𝜃𝐽(𝜃; 𝑥 𝑖:𝑖+𝑛 ; 𝑦 (𝑖:𝑖+𝑛) )

За каждую эпоху отбирается 𝒏 данных (𝒏 = 𝟓𝟎 − 𝟐𝟓𝟎) небольшой набор обучающих данных из всего набора данных (выборки - batch). Применение этого метода дает следующие преимущества:

1. Уменьшает дисперсию разброса параметров что ведет к более устойчивой сходимости.
2. Позволяет использовать матричные и тензорные операции, входящие в состав современных библиотек методов глубокого обучения Все это делает использование SGD на минибатчах очень эффективными для решения задач по глубокому обучению.
   1. **NAG. (Nesterov accelerated gradient)**

𝑣𝑡 = 𝑣𝑡−1 + 𝜂 ∇𝜃𝐽(𝜃 –𝛾𝑣𝑡−1)

𝜃 = 𝜃 –𝑣

* 1. **Adagrad. (Adapted gradient)**

Оптимизатор первого порядка. Скорость обучения не постоянна, а зависит от целевой функции. На каждой итерации глобальная скорость обучения делится на

𝒍𝟐 −норму прошлых градиентов вплоть до текущей.

**Adagrad. (Adapted gradient)**

* 1. **Adam. (Adaptive moment estimation)**

Описание методов оптимизации можно найти в INTERNETе и в моей лекции № 4.

**Adam. (Adaptive moment estimation)**

*.*

*)*

*)*

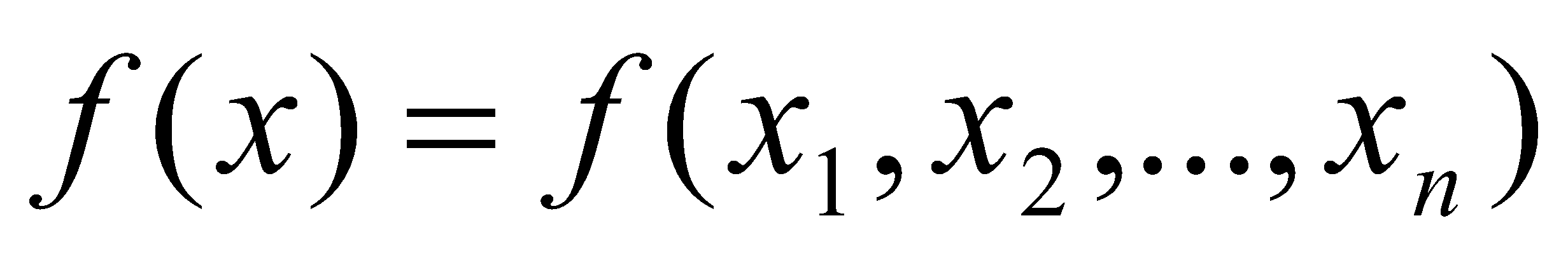
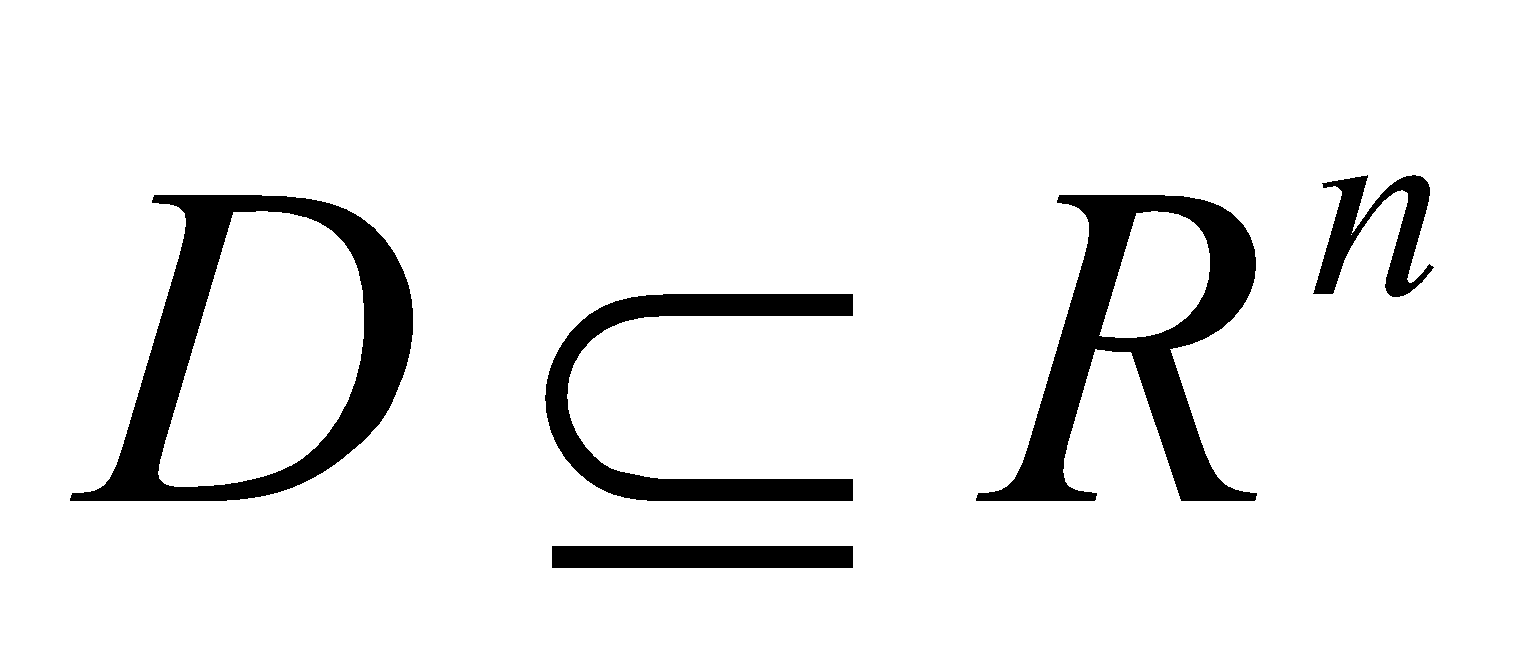
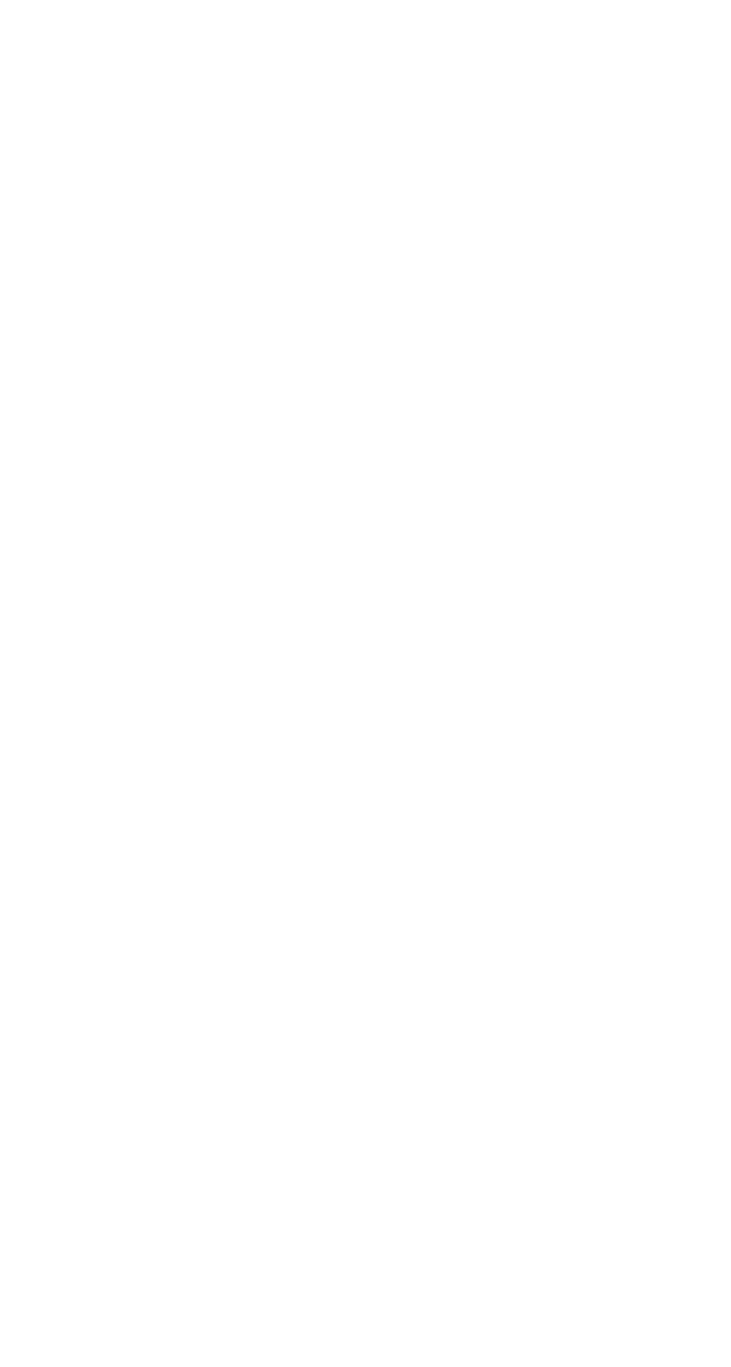
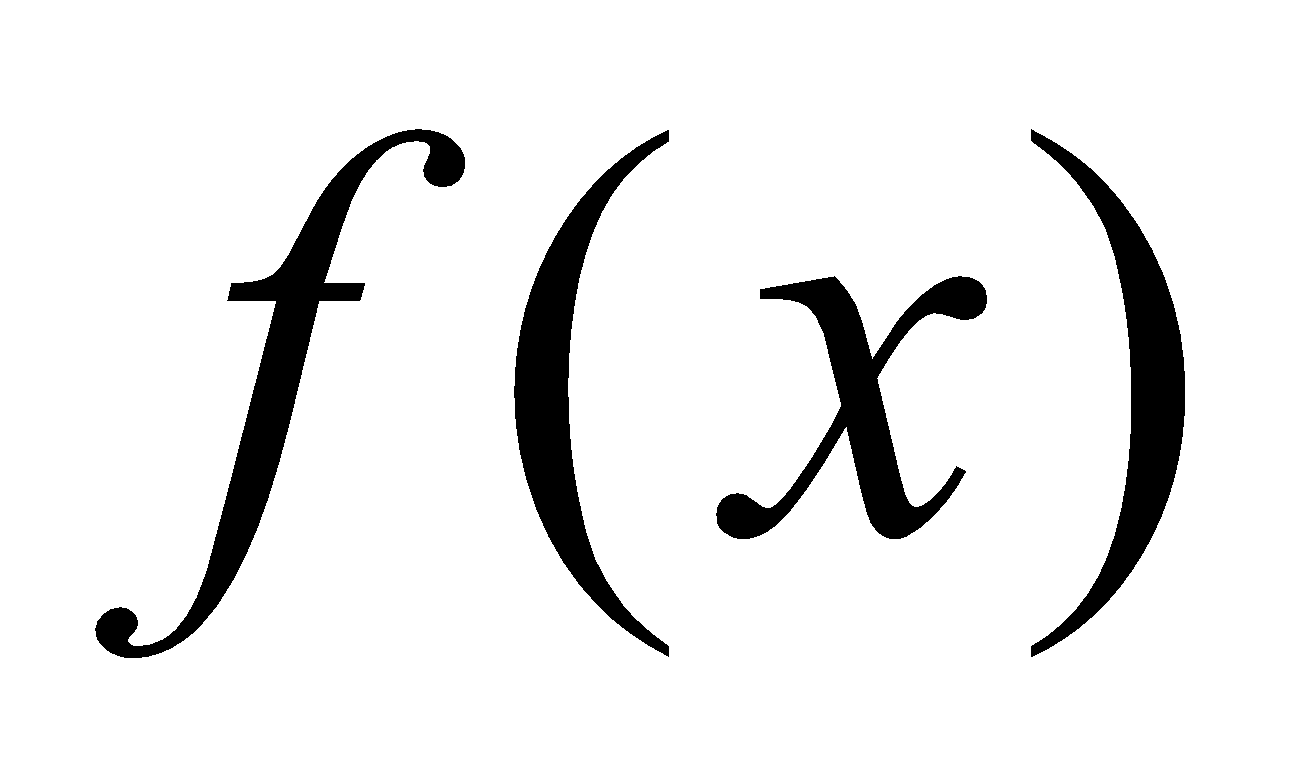
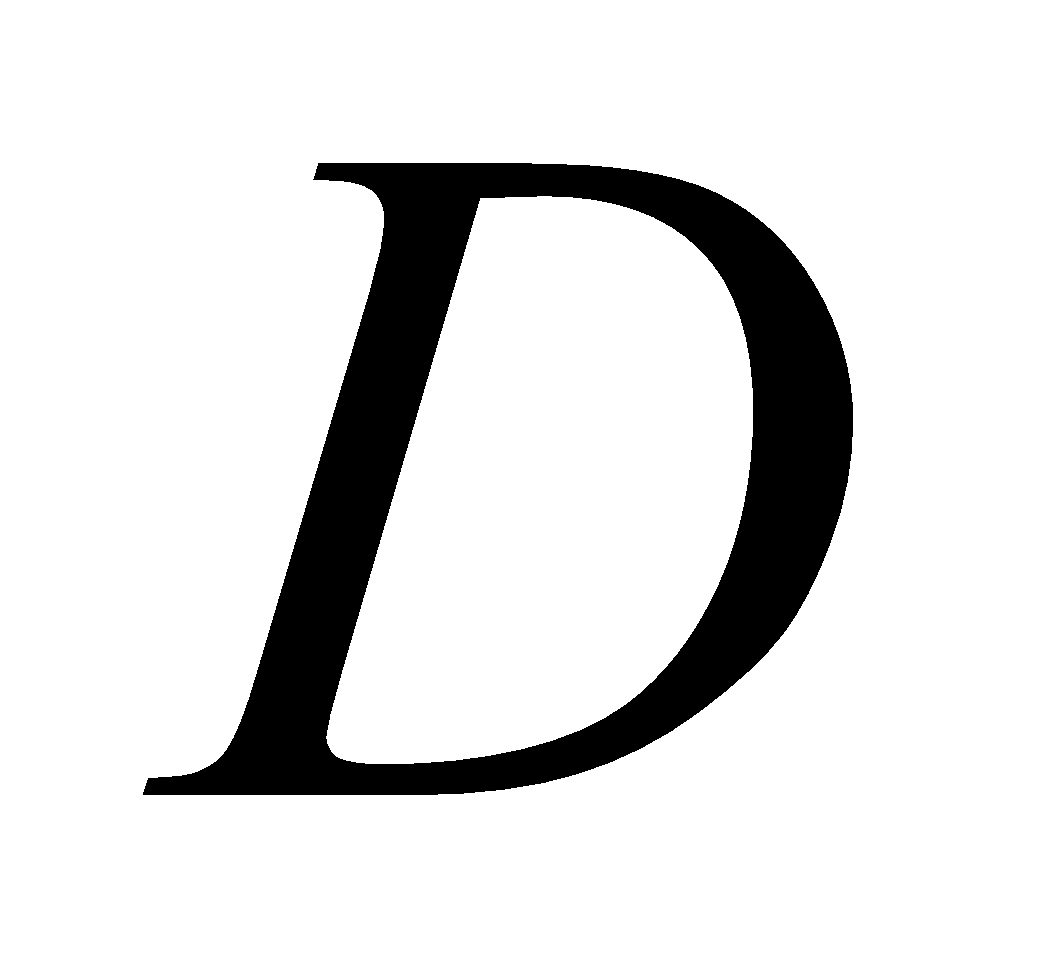
.

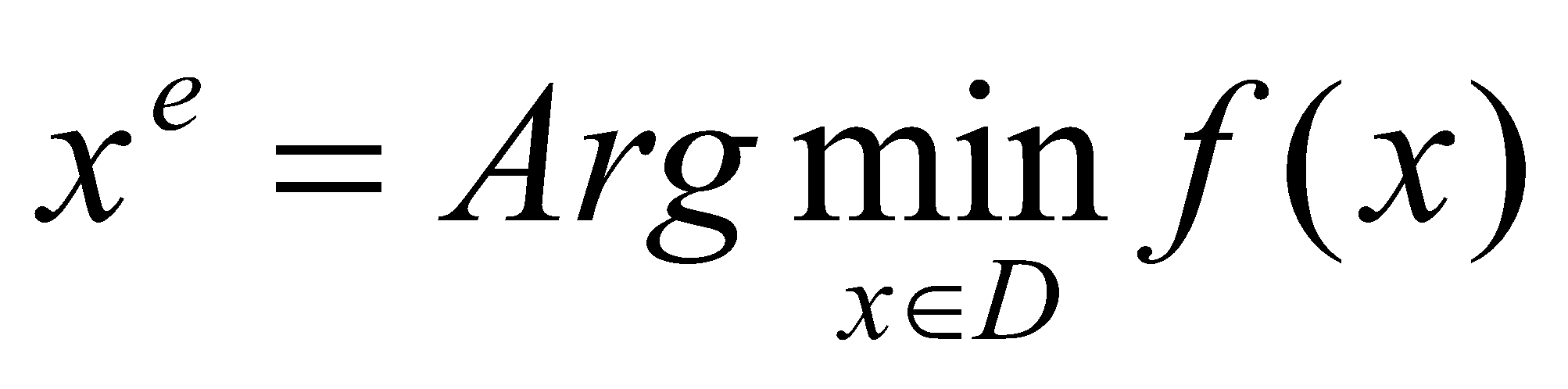
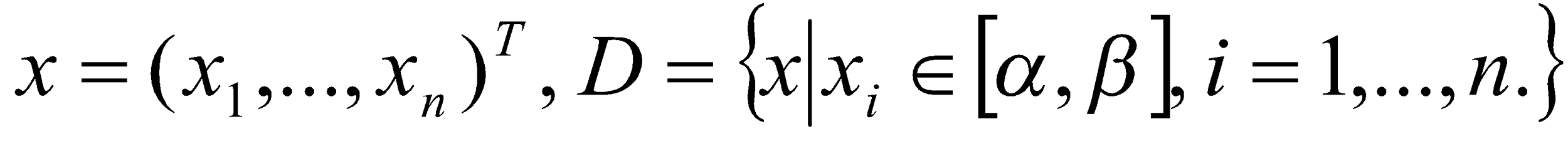
.

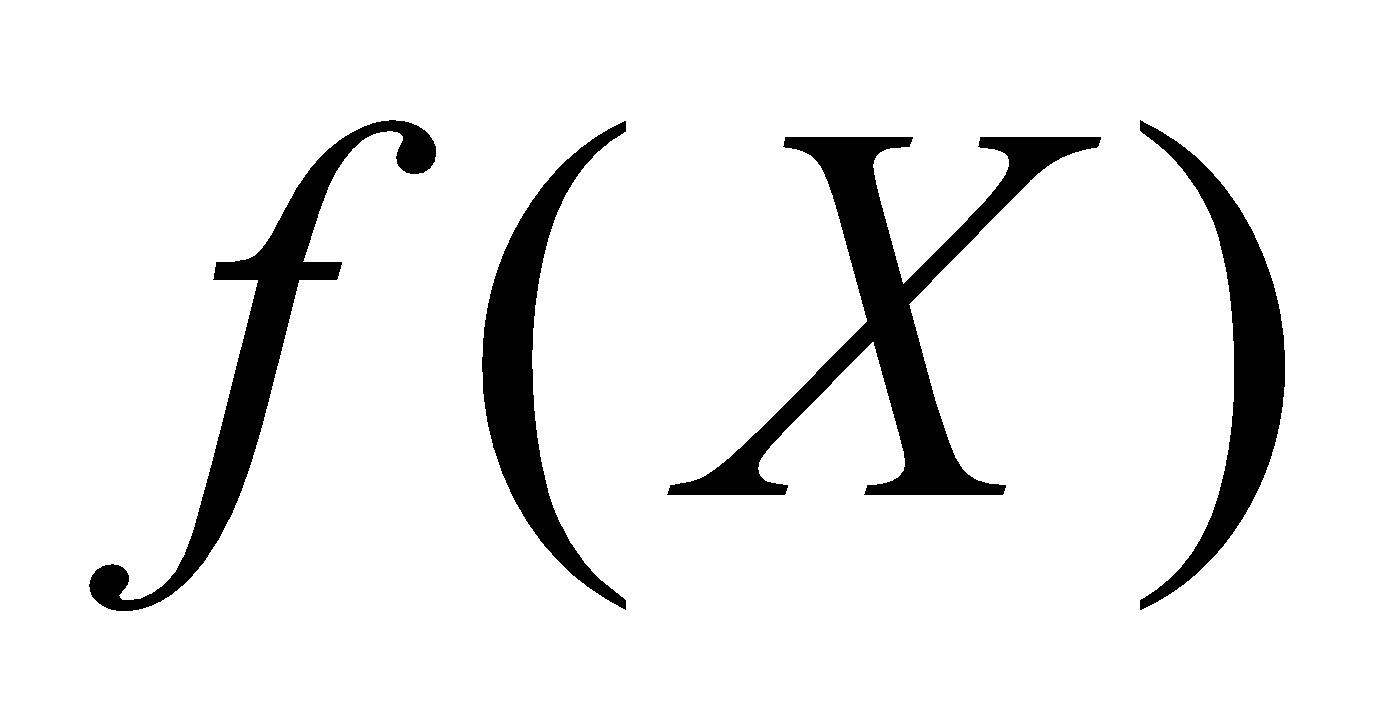
.

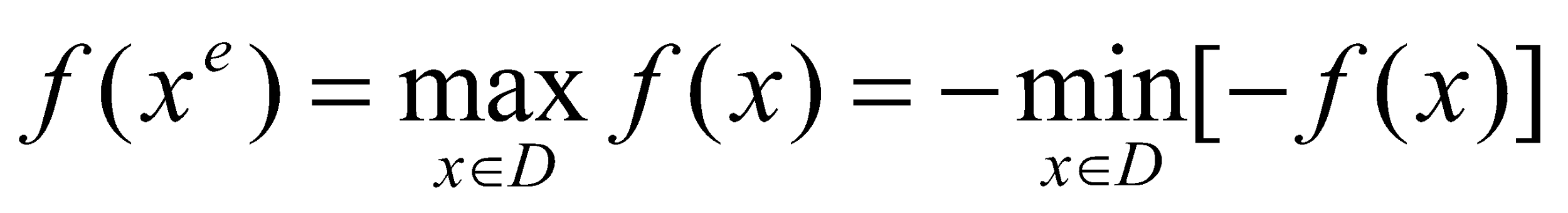
1. **Генетический алгоритм.**

**Постановка задачи**

Дана целевая функция , определенная на множестве допустимых решений. Требуется найти глобальные минимумы заданных функций  на допустимом множестве . То есть такую точку

, где . **(1)**

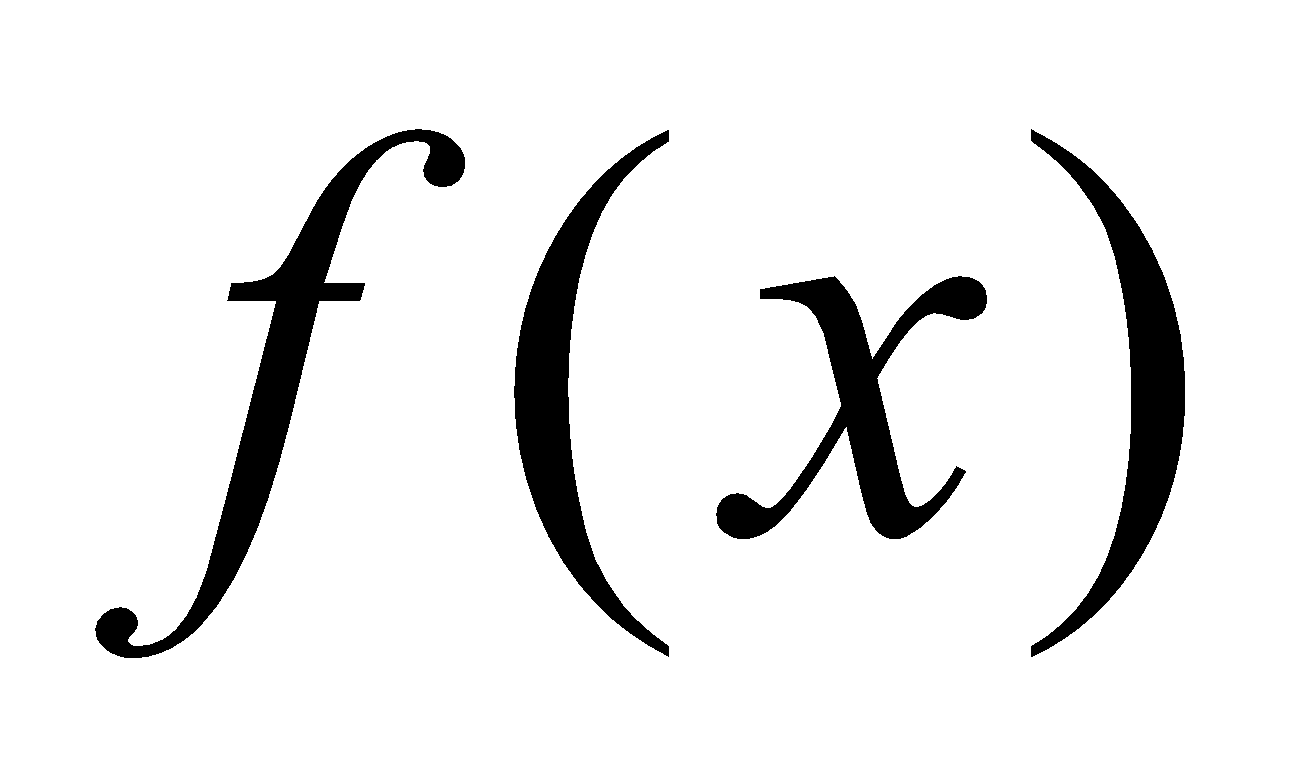
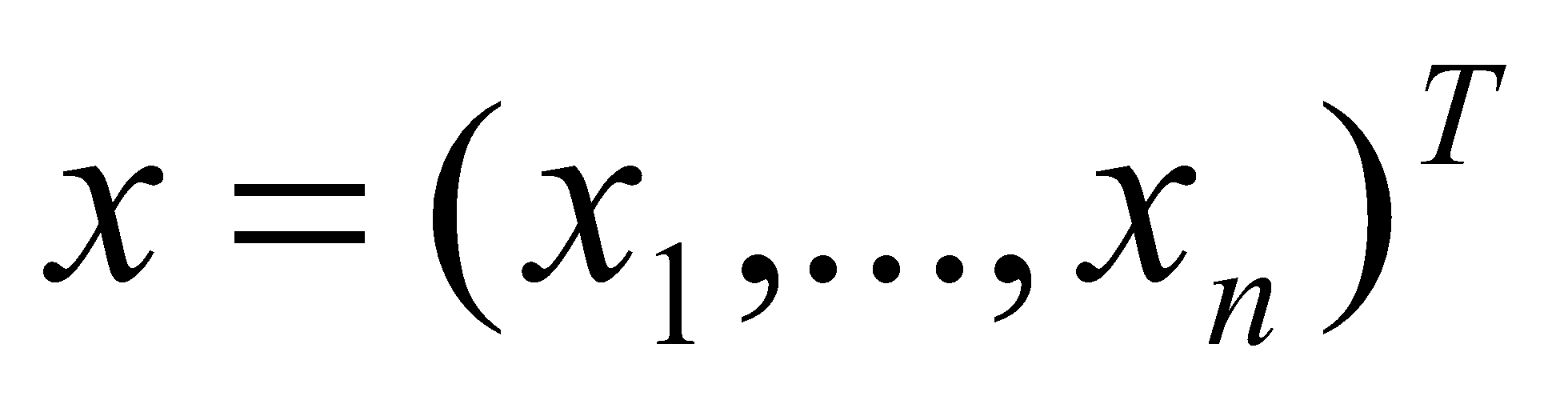
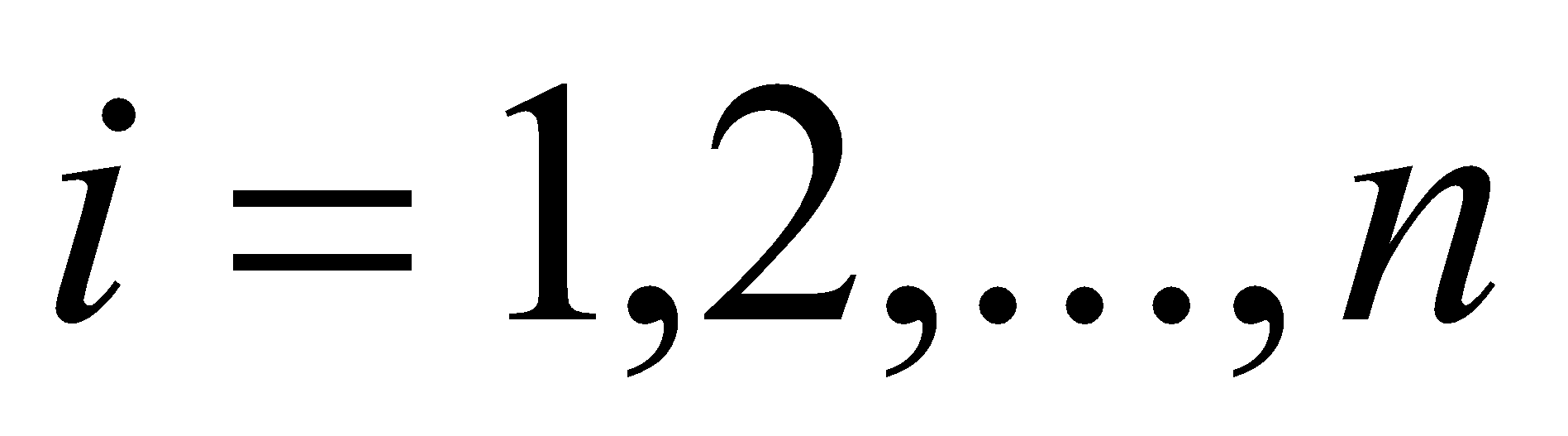
Задача поиска максимума целевой функции  сводится к задаче поиска минимума путем замены знака перед функцией на противоположный:

**.**

* 1. **Стратегия поиска**

Генетические алгоритмы имитируют природные способы оптимизации, присущие процессам эволюции живых систем. А именно:

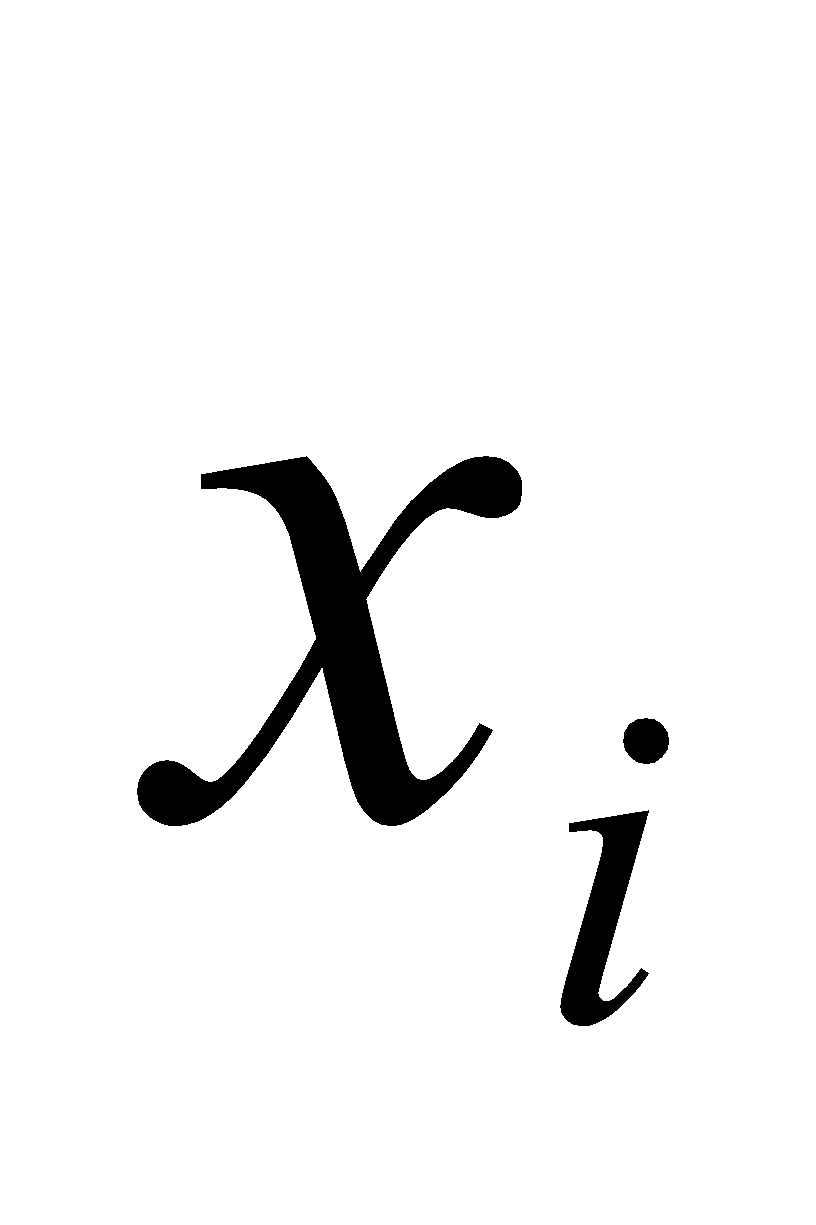
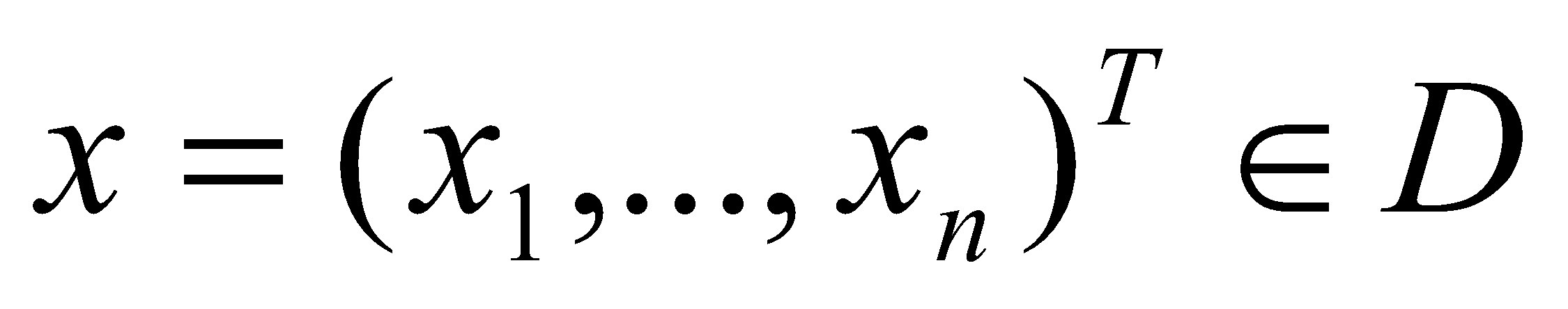
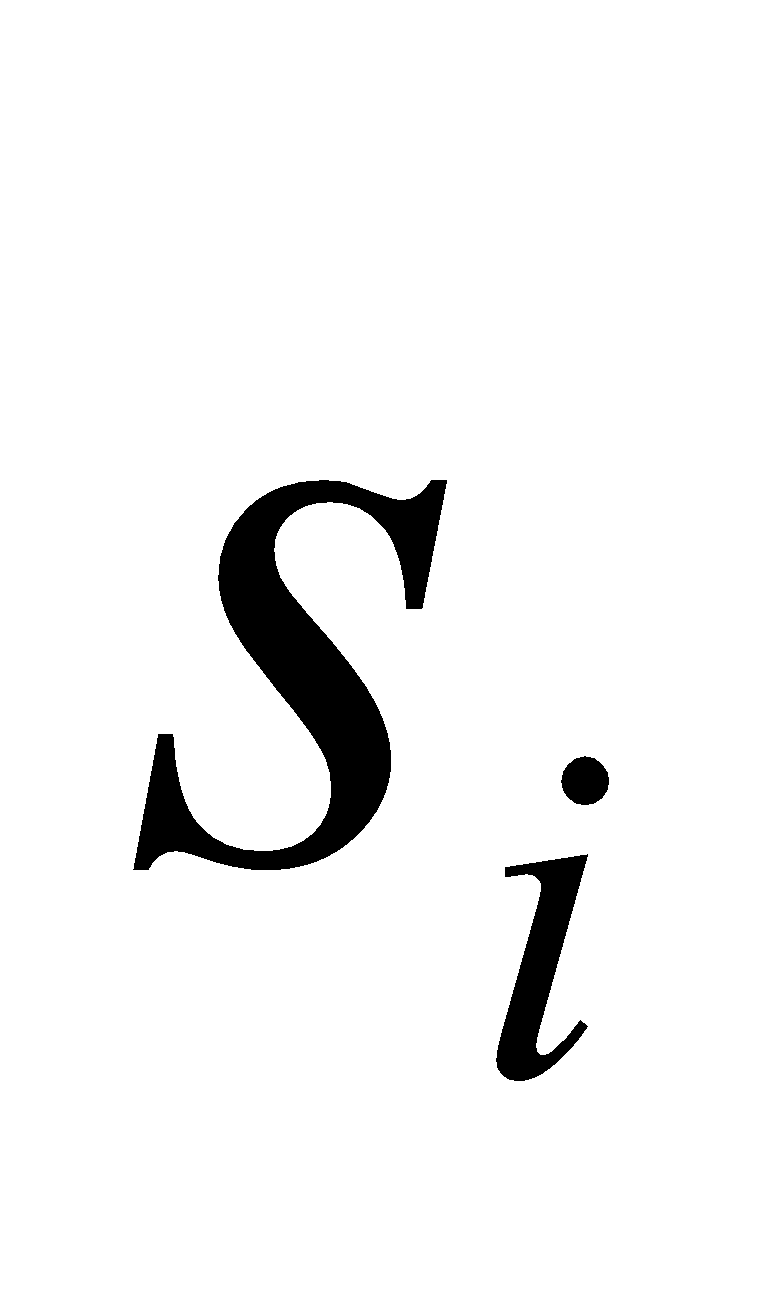
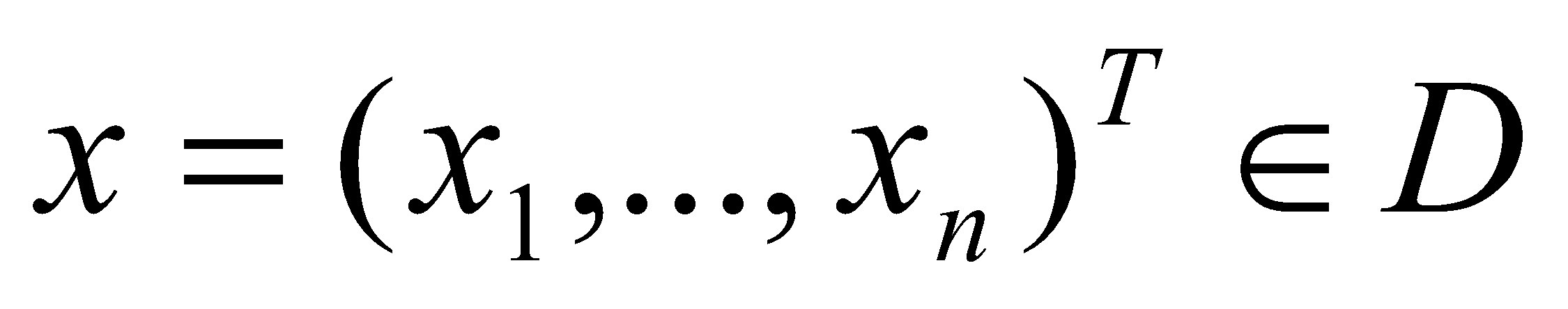
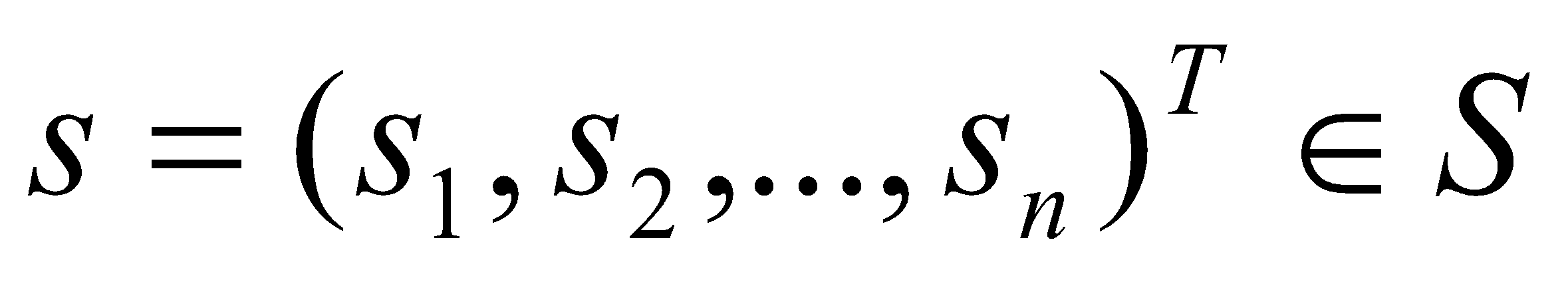
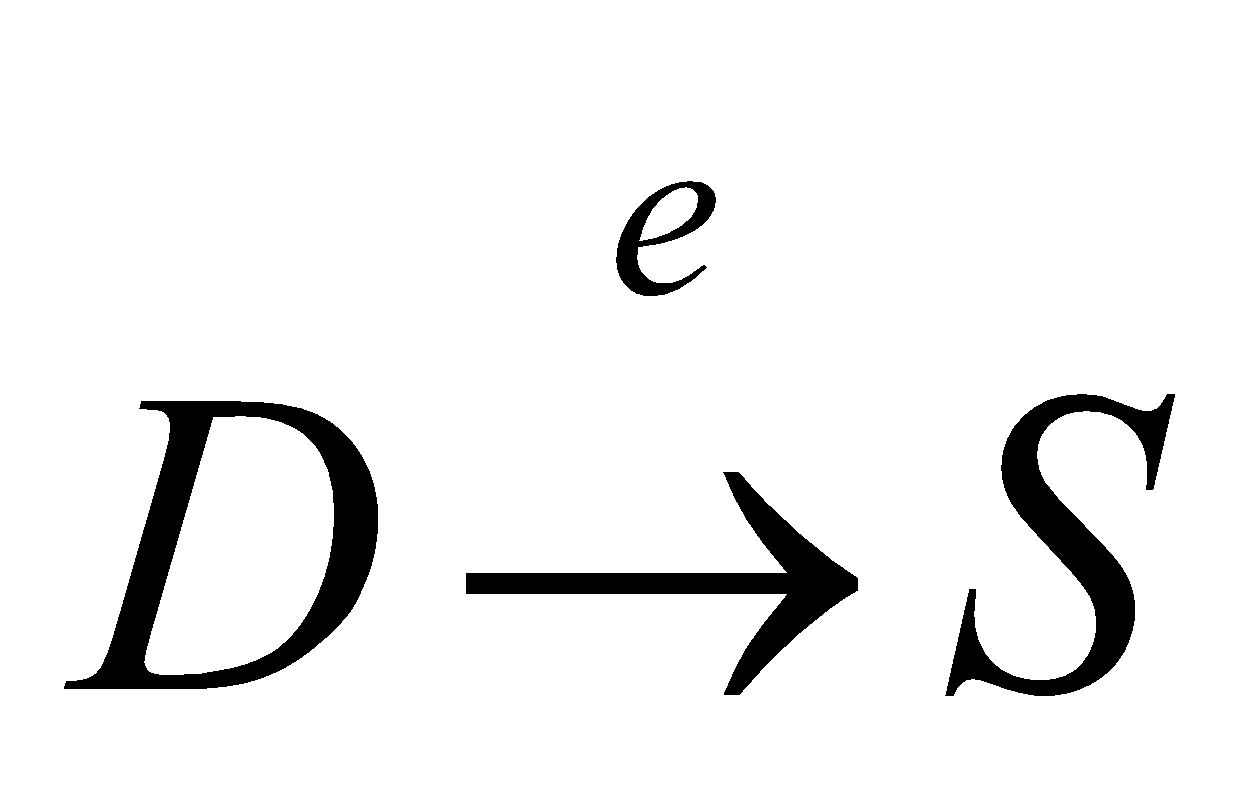
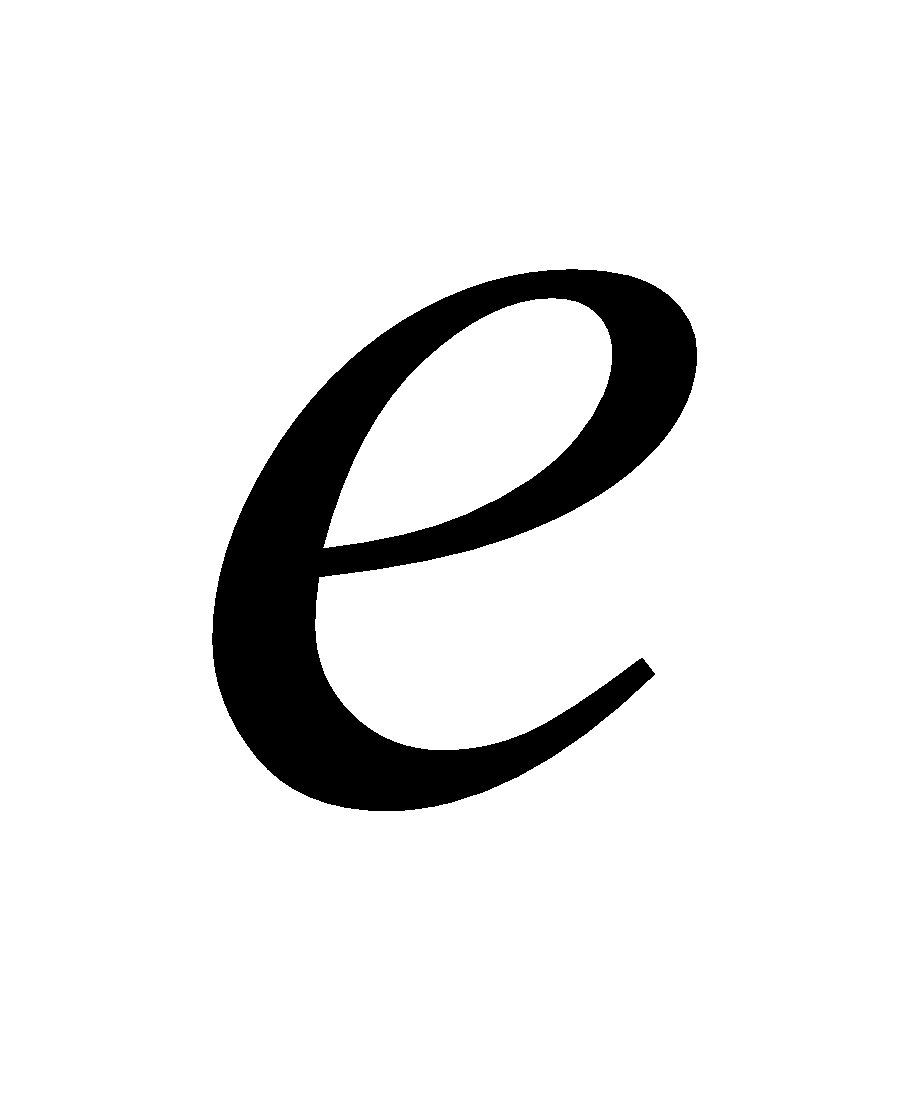
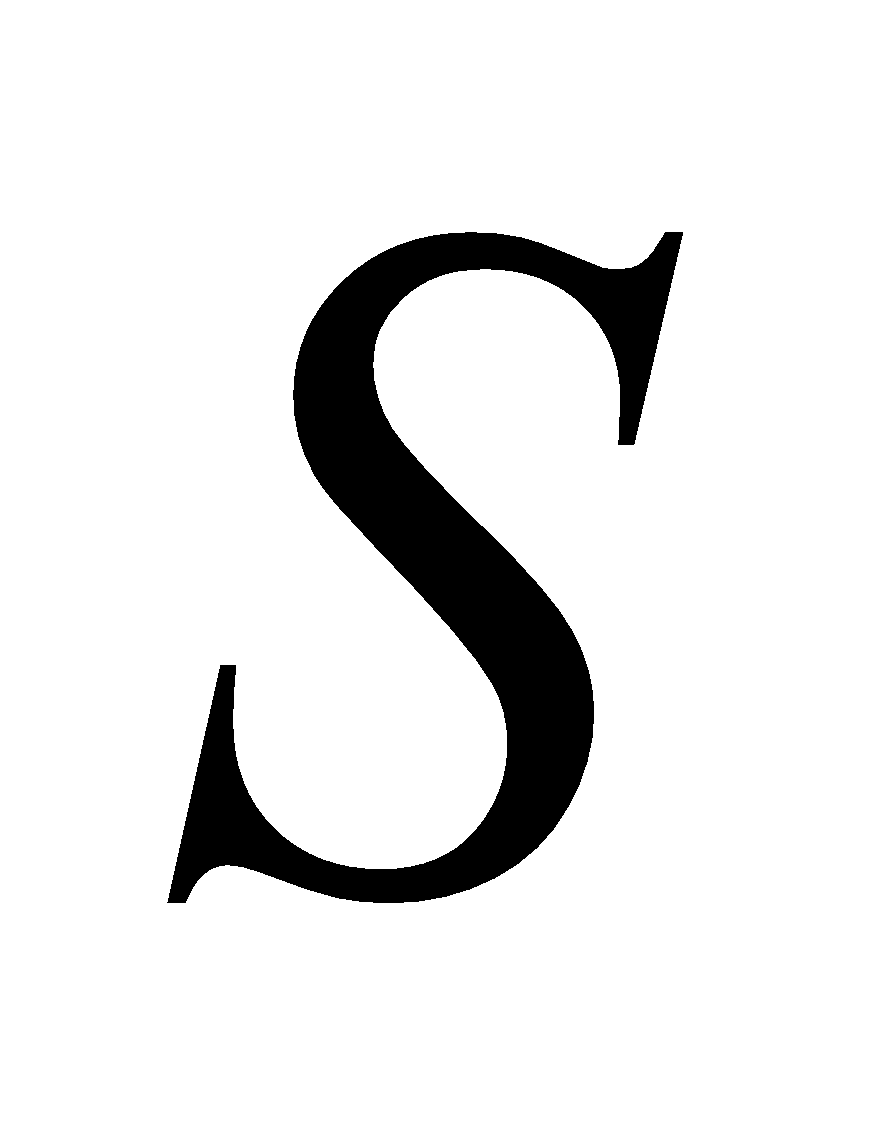
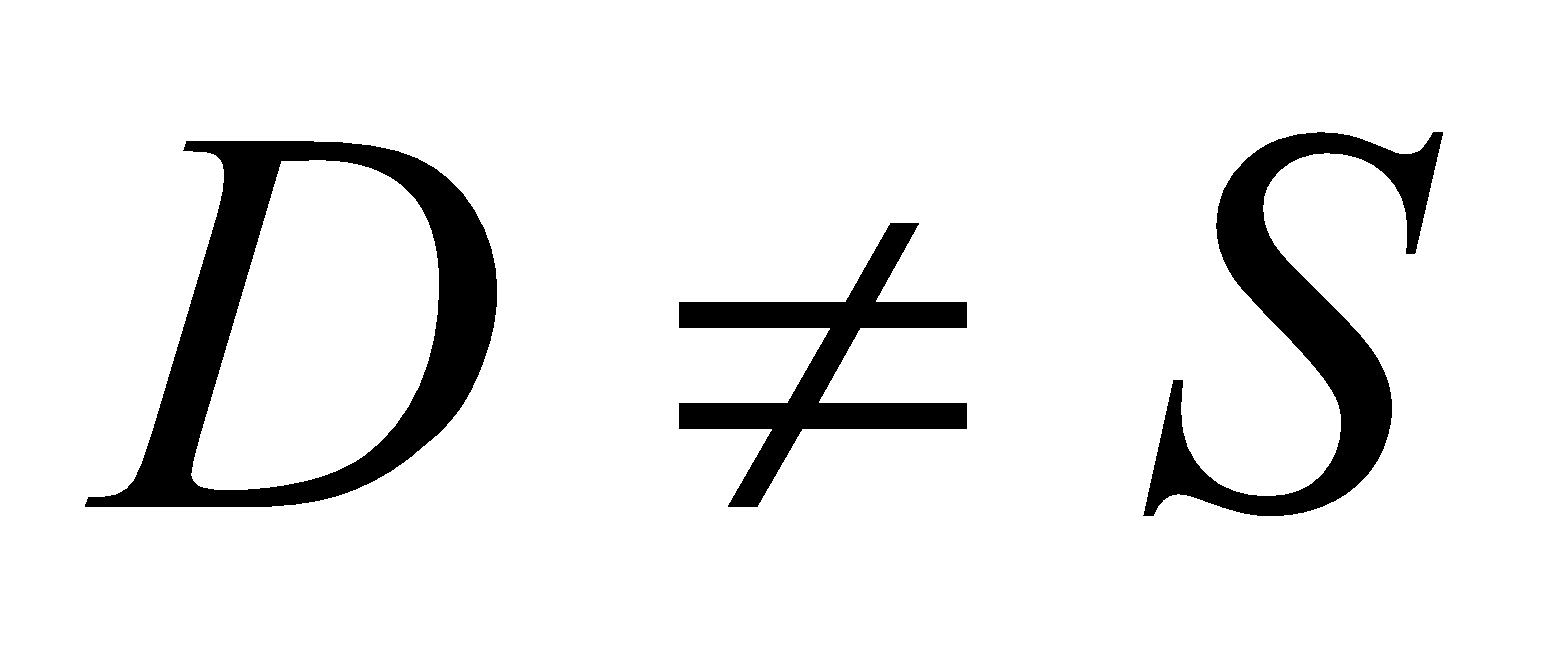
* генетическое наследование;
* изменчивость;
* естественный отбор.

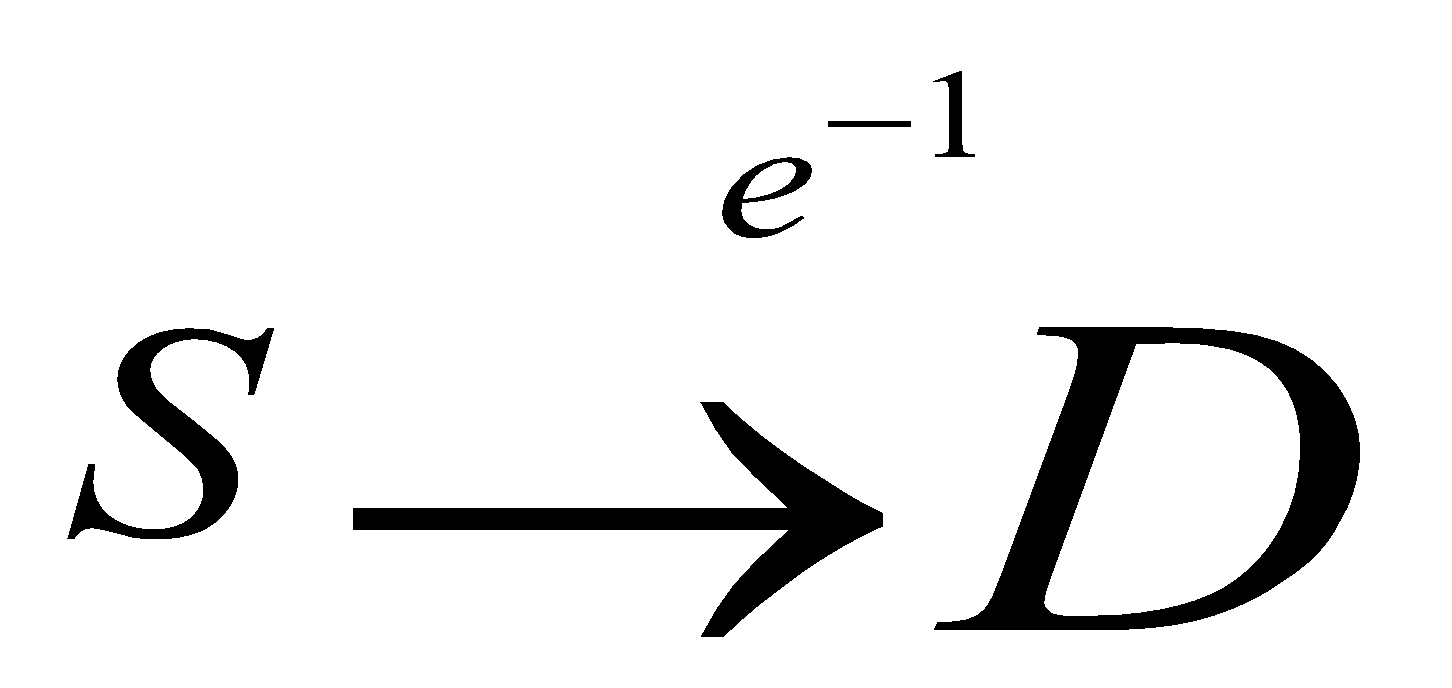
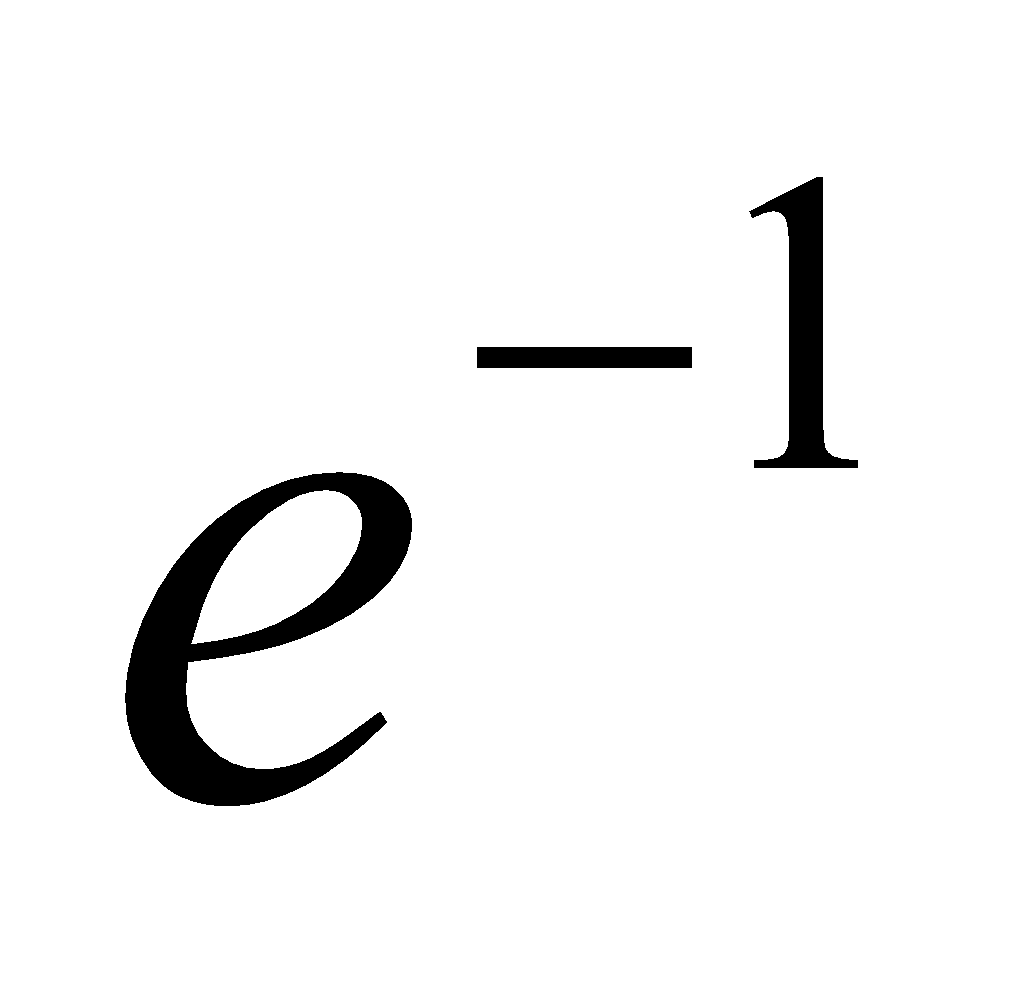
***Целевая функция***  соответствует природному понятию ***приспособленности*** живого организма. Вектор переменных  целевой функции называется ***фенотипо*м**, а отдельные его параметры – ***признаками*** .

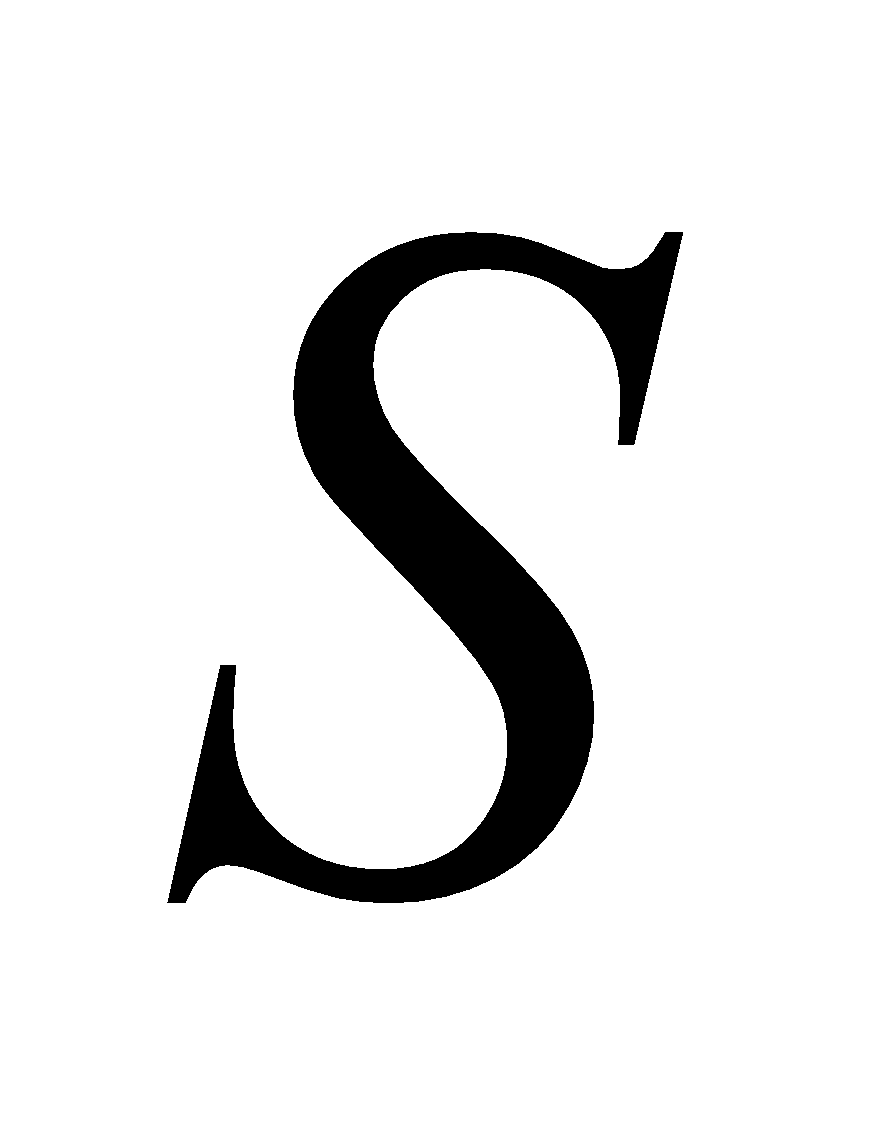
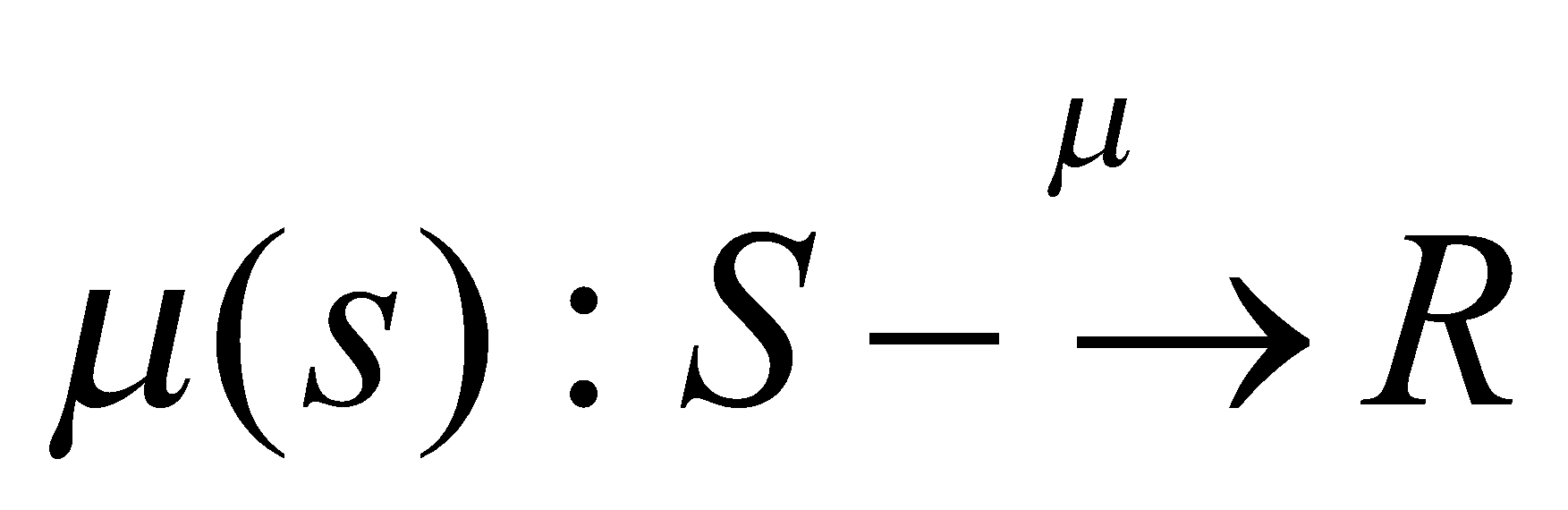
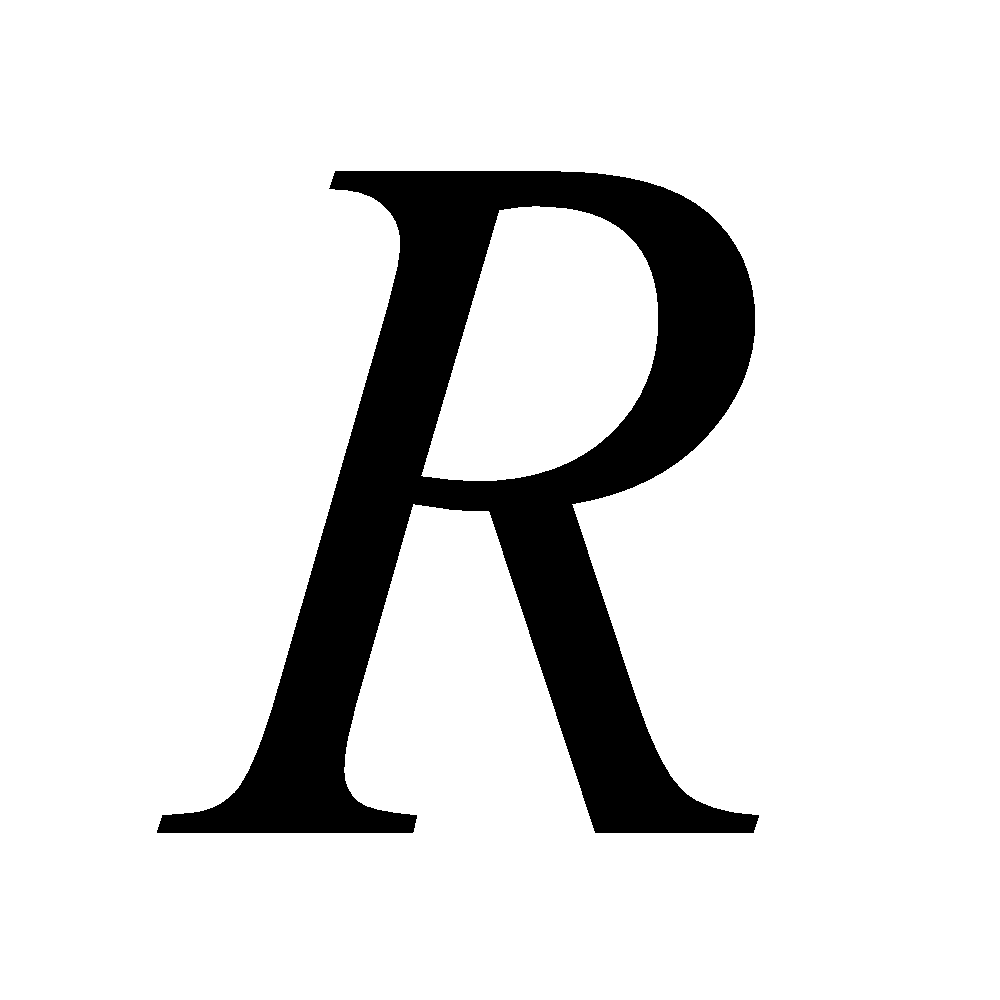
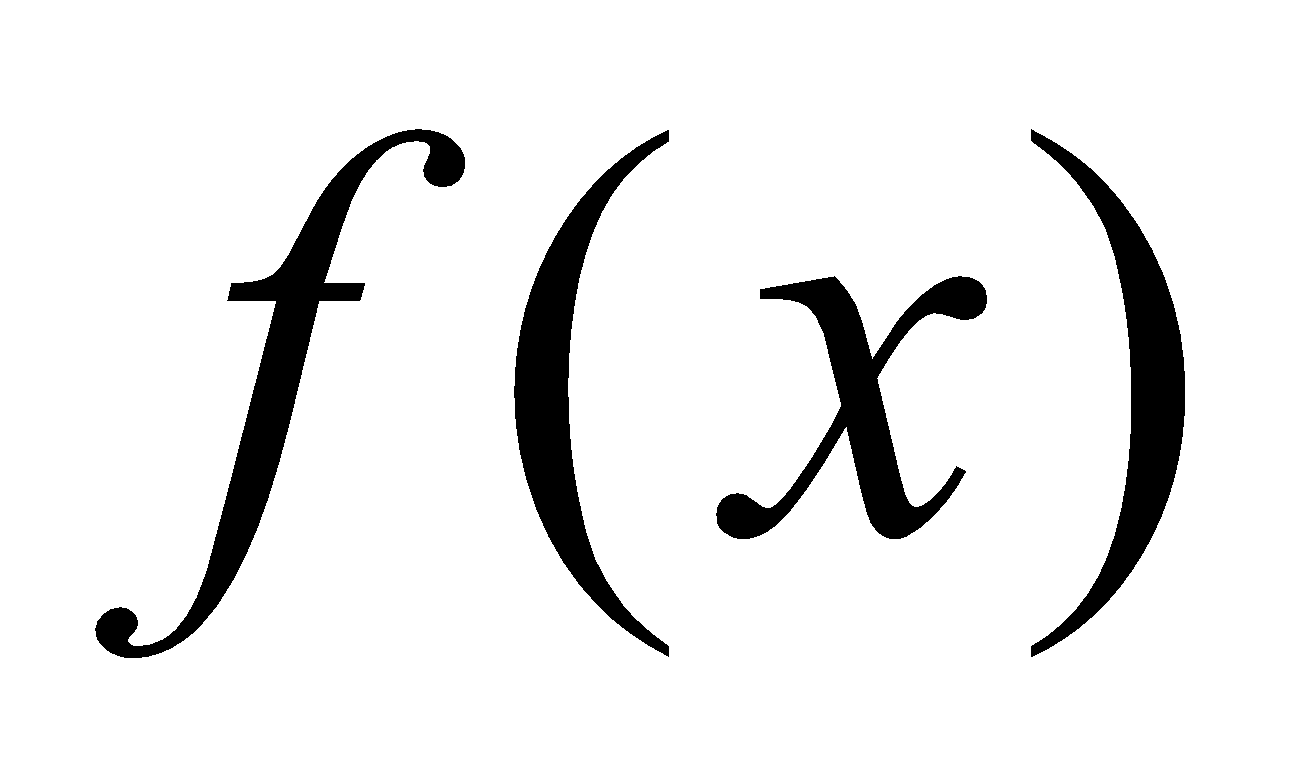
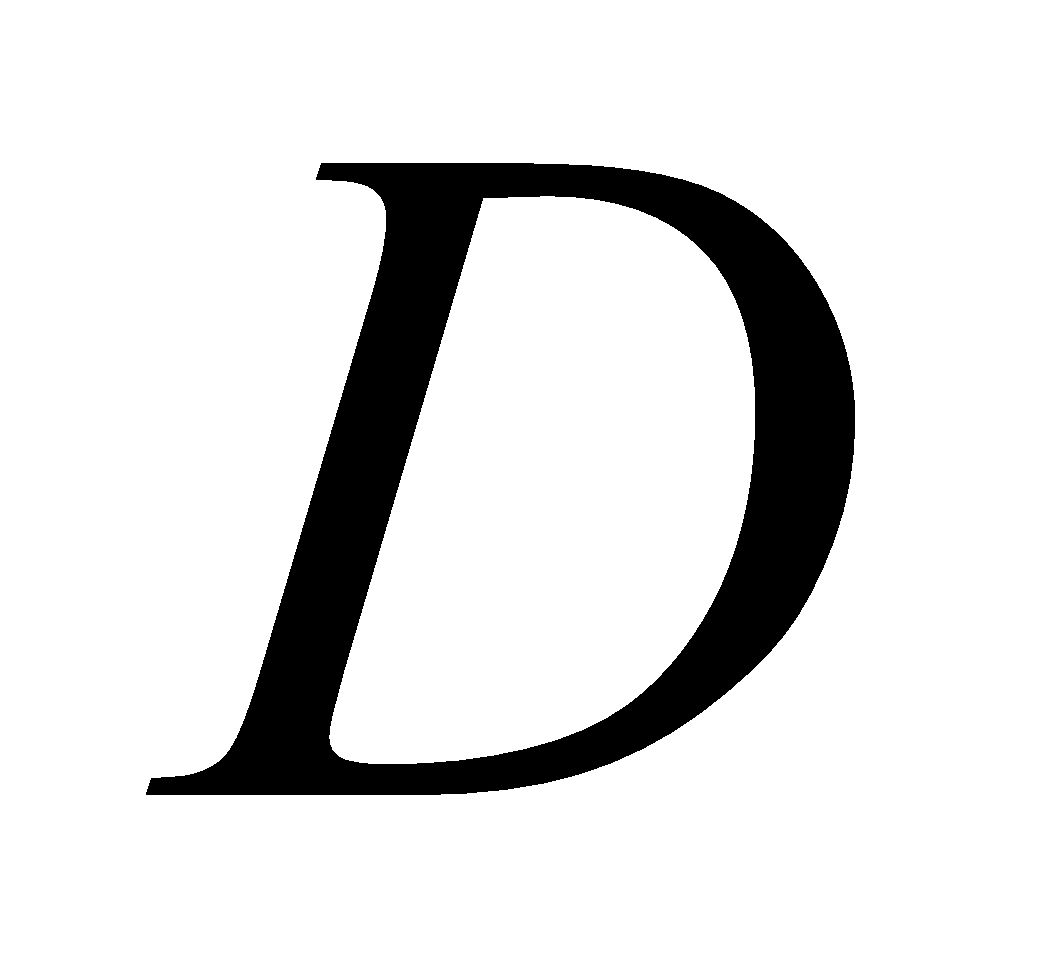
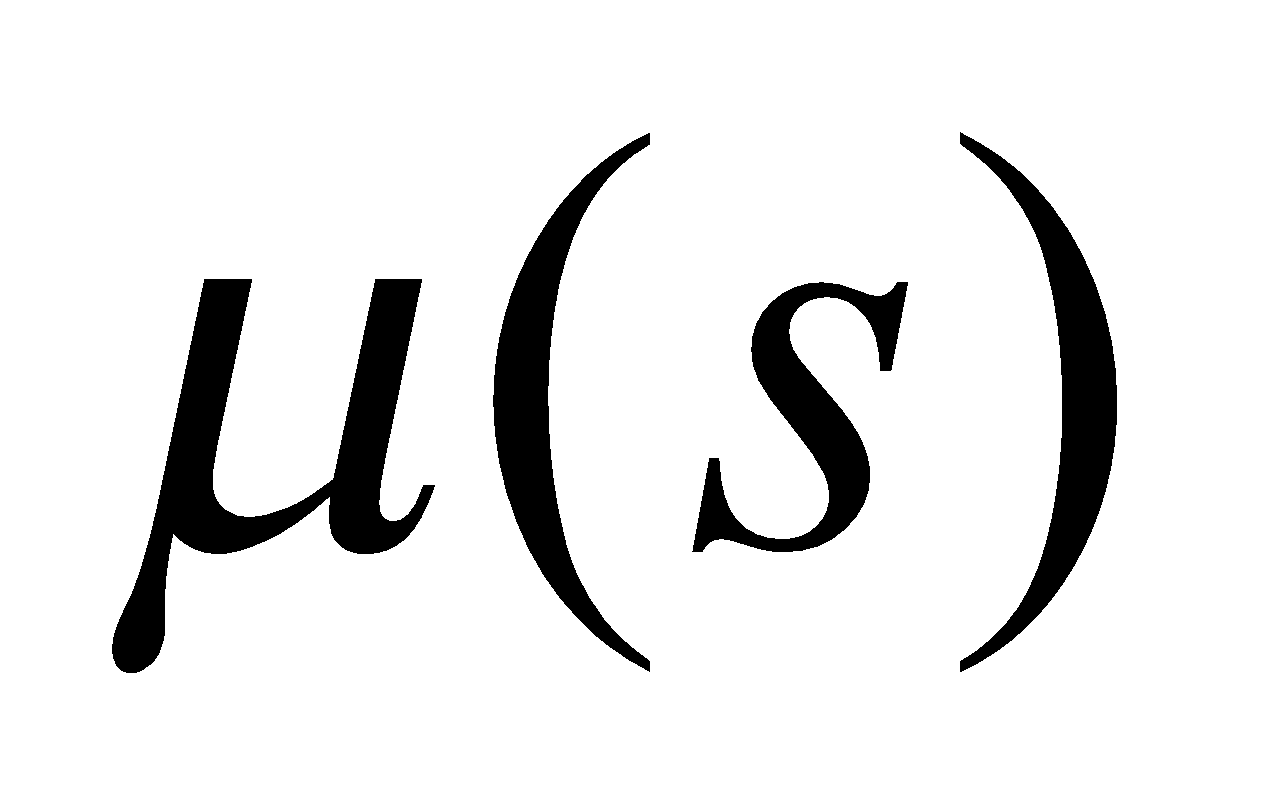
Любой живой организм может быть представлен своим генотипом и фенотипом.

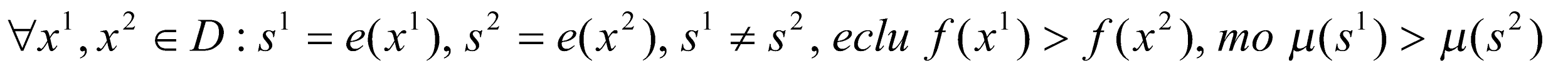
***Генотип***– это совокупность наследственных признаков, информация о которых заключена в хромосомном наборе генов.

***Фенотип***– совокупность всех признаков и свойств организма, формирующихся в процессе взаимодействия его генотипа и внешней среды. Каждый ген имеет своё отражение в фенотипе.

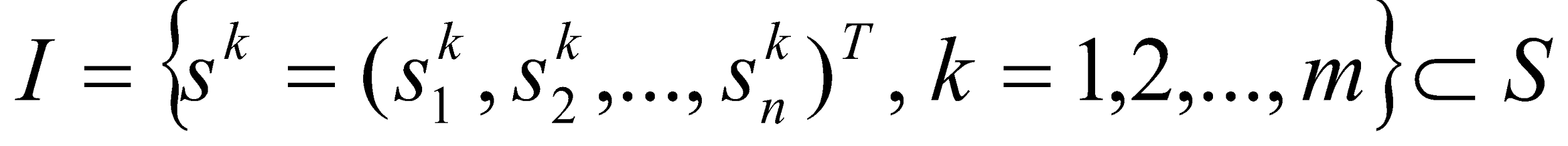
***Генетические алгоритмы***ведут поиск решения на уровне генотипа. Каждую координату  вектора  представляют в некоторой форме , удобной для использования в генетическом алгоритме и называется *геном*. Для этого необходимо выполнить преобразование, в общем случае не взаимно однозначное, вектора переменных  в некоторую структуру , называемую ***хромосомой******(генотипом, особью****): ,* где  функция кодирования, - пространство представлений (как правило )*.*

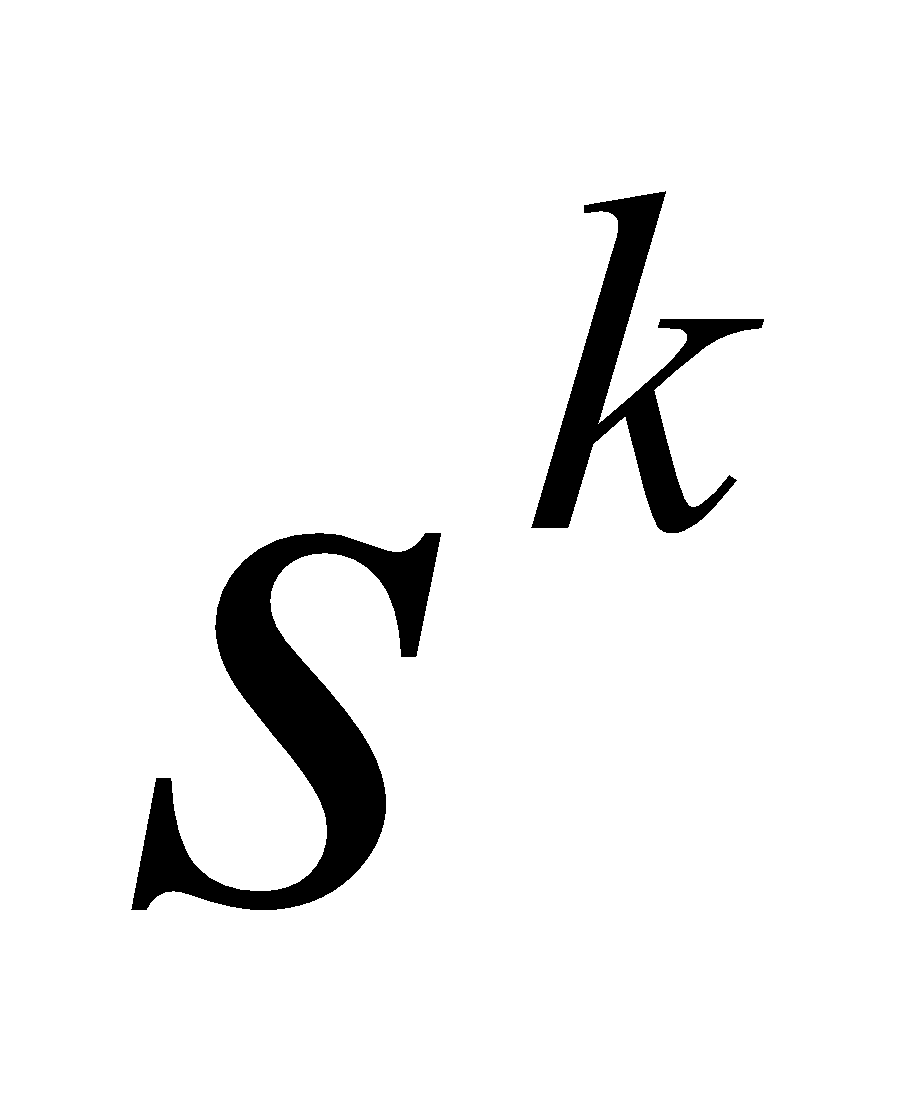
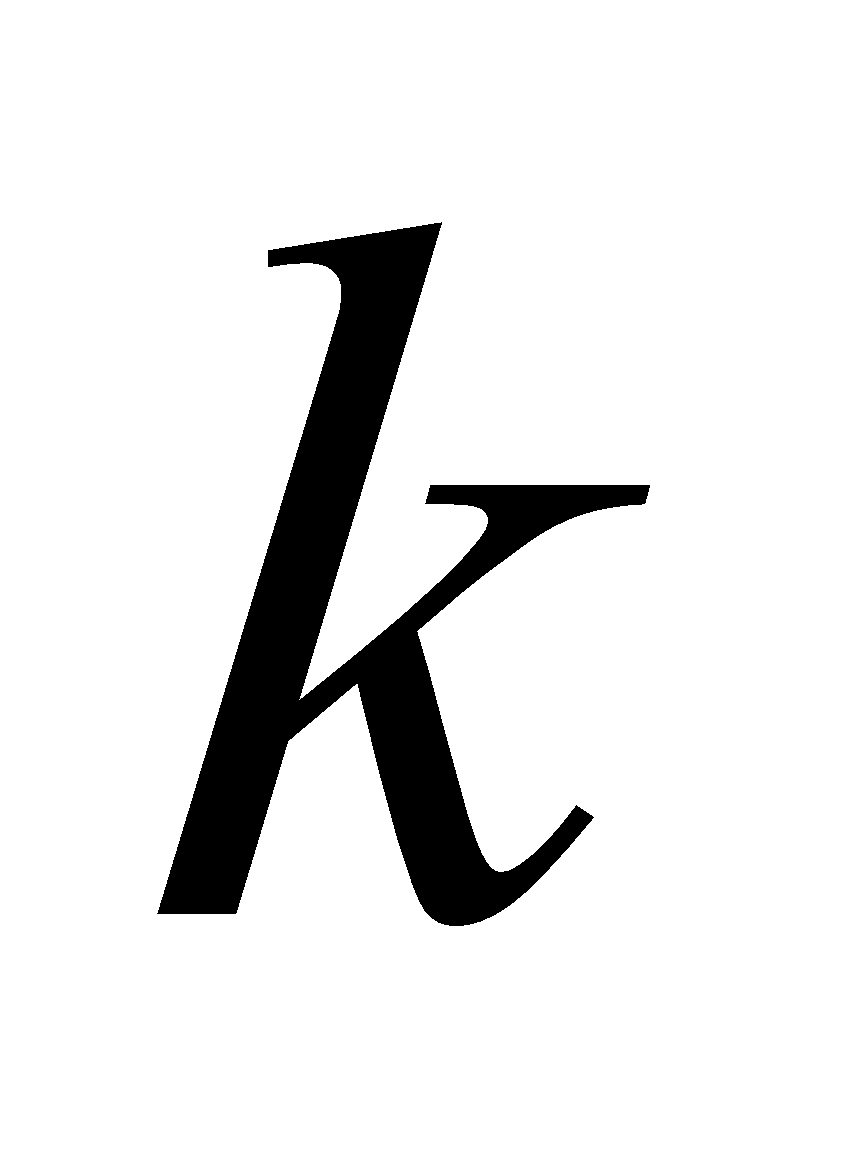
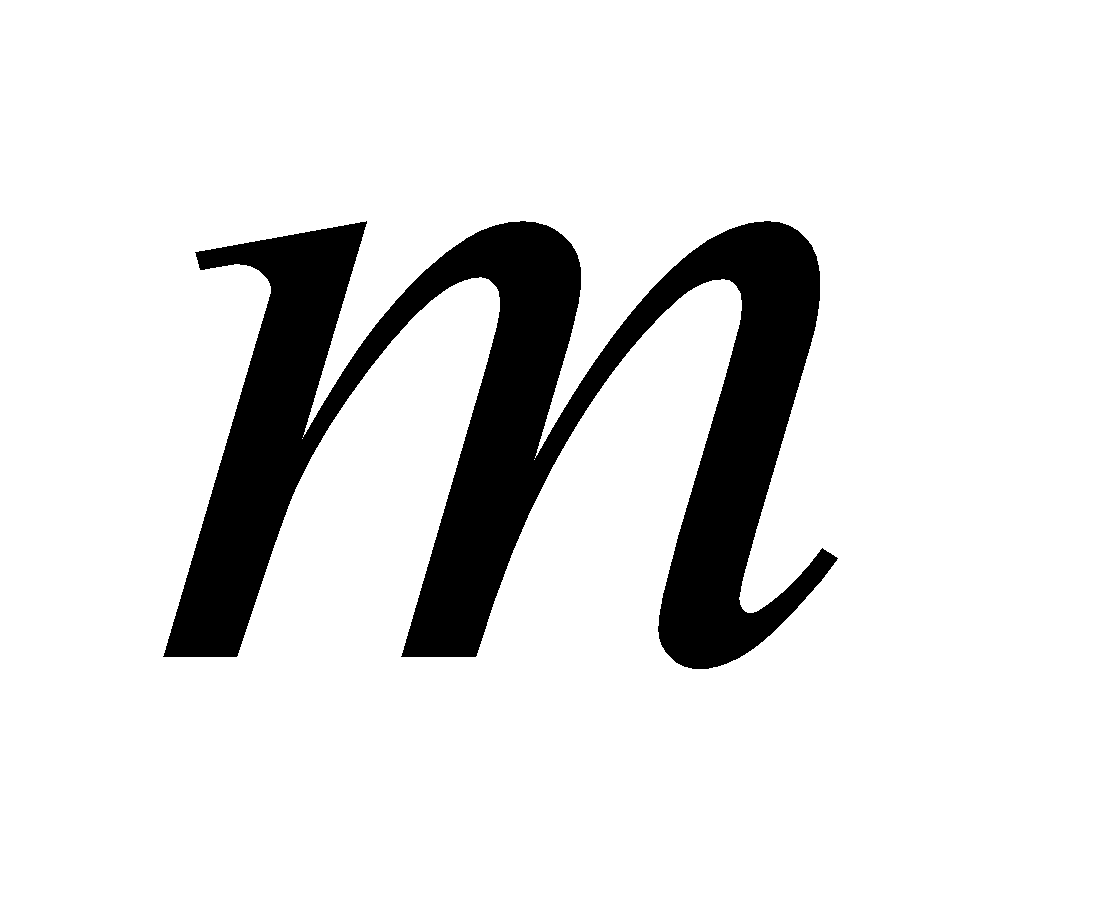
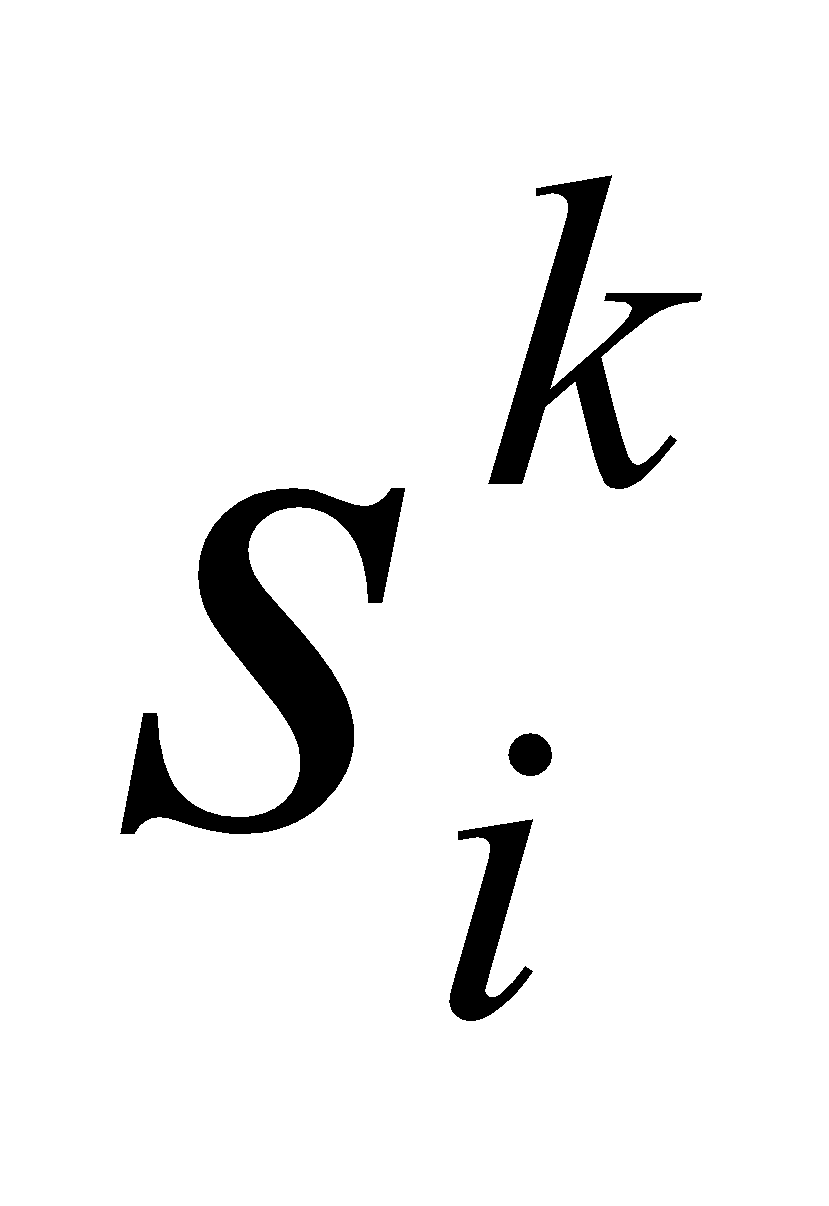
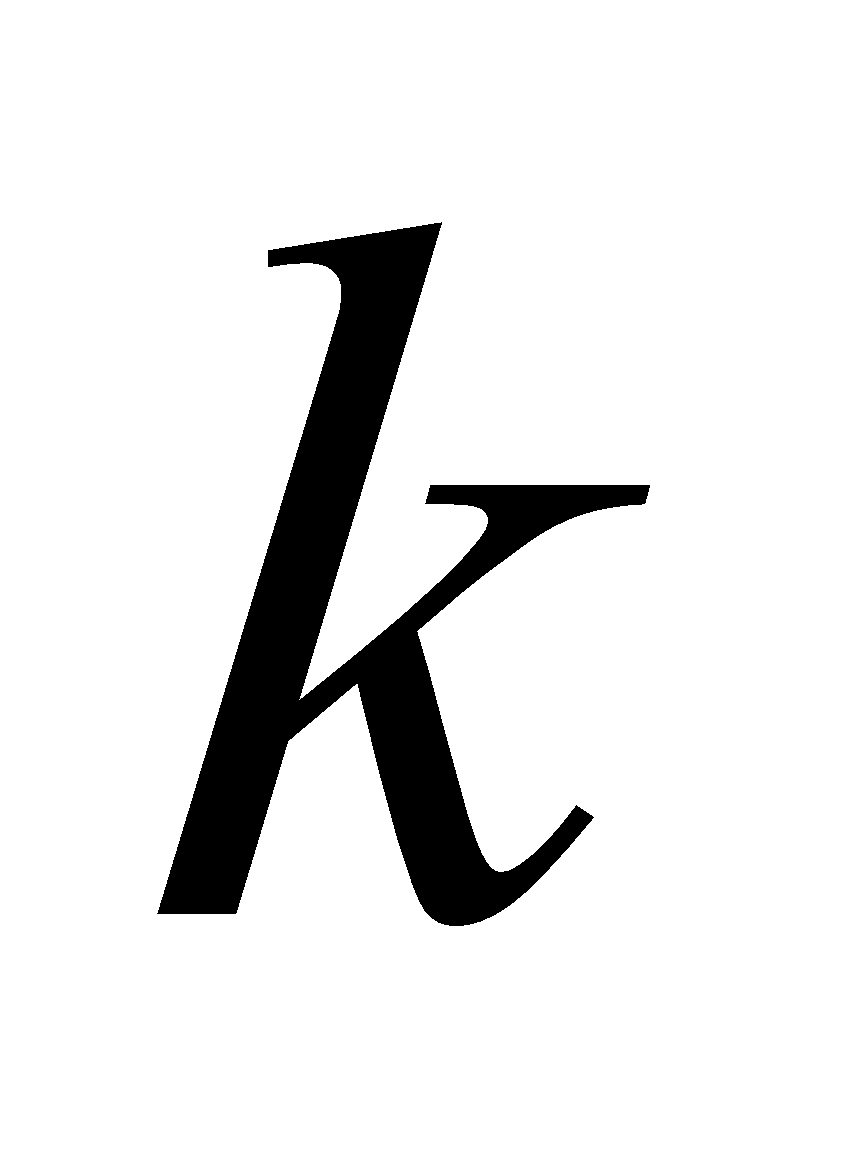
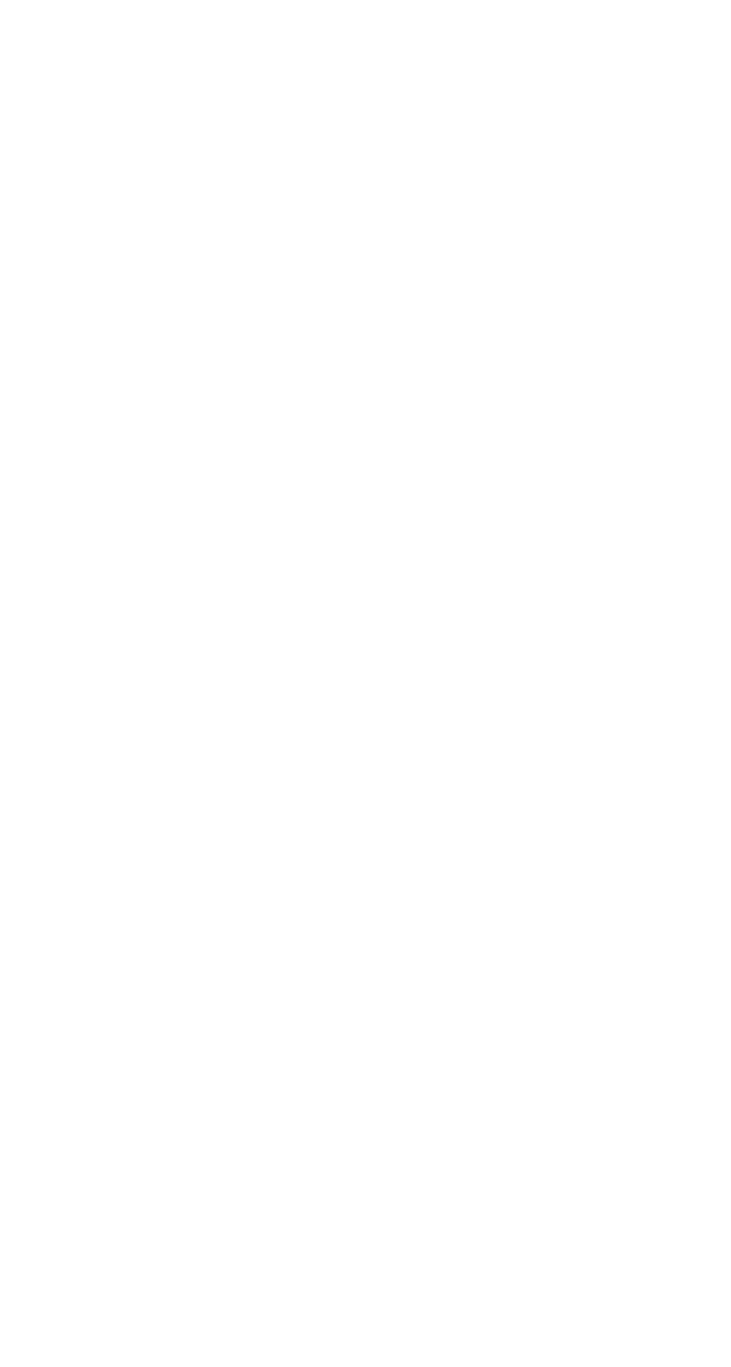
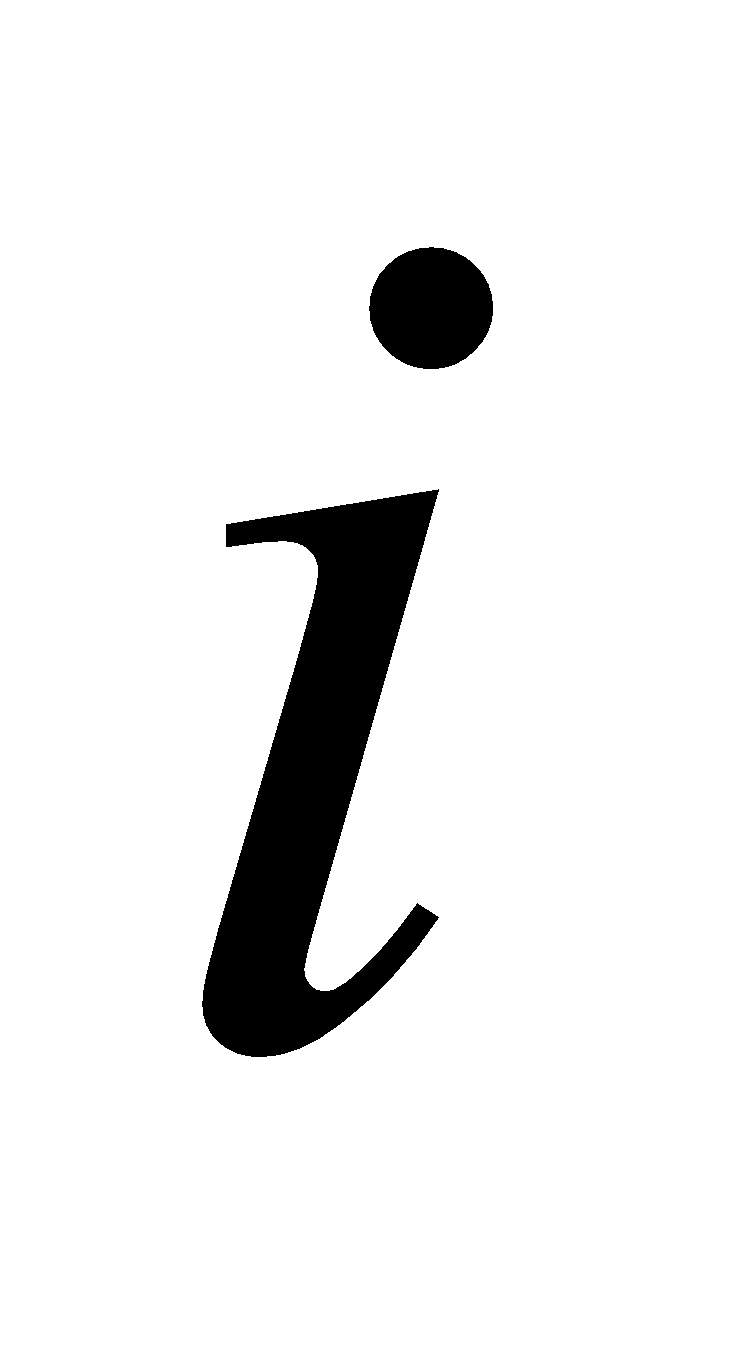
Для того, чтобы восстанавливать по хромосоме решение, необходимо задать обратное преобразование * ,* где  - функция декодирования.

В пространстве представлений  вводится так называемая ***функция приспособленности (функция фитнеса)*** *,* где  - множество вещественных чисел, аналогичная целевой функции  на множестве . Функция  может быть любая функция, удовлетворяющая условию:

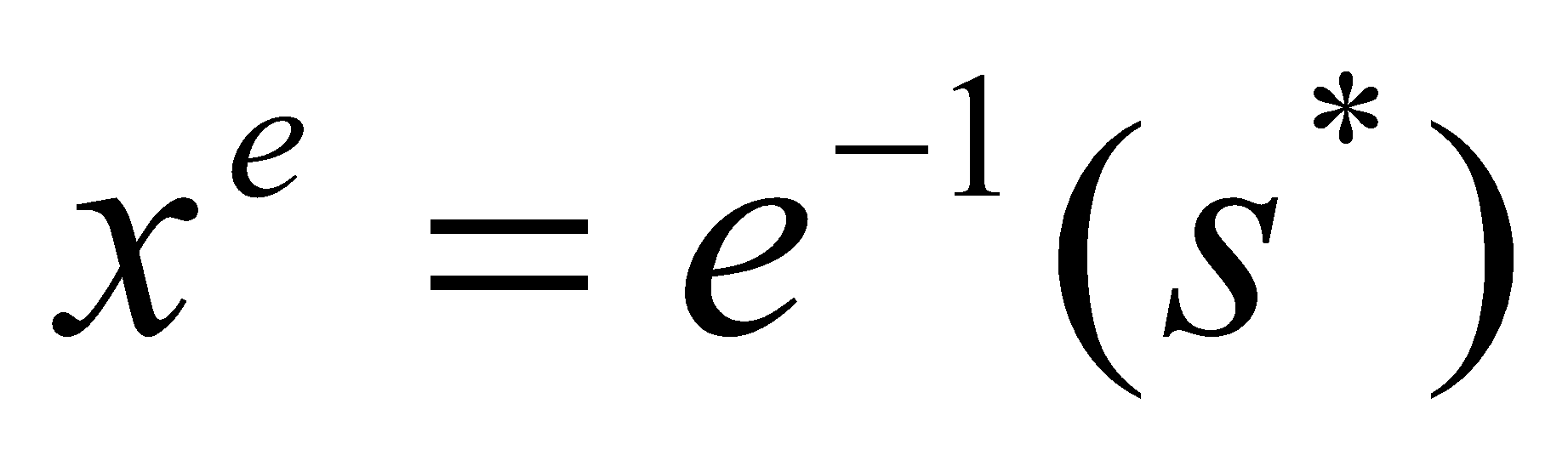
.

При решении используются конечные наборы:



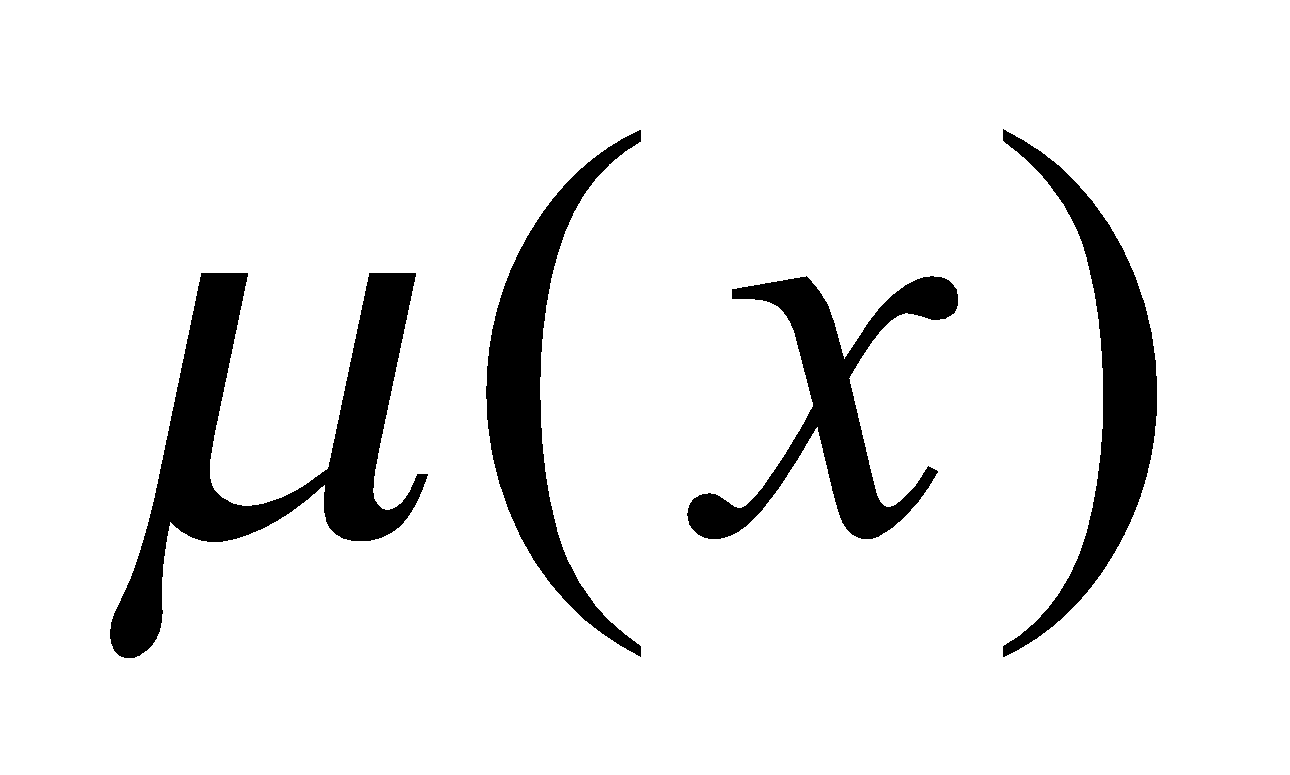
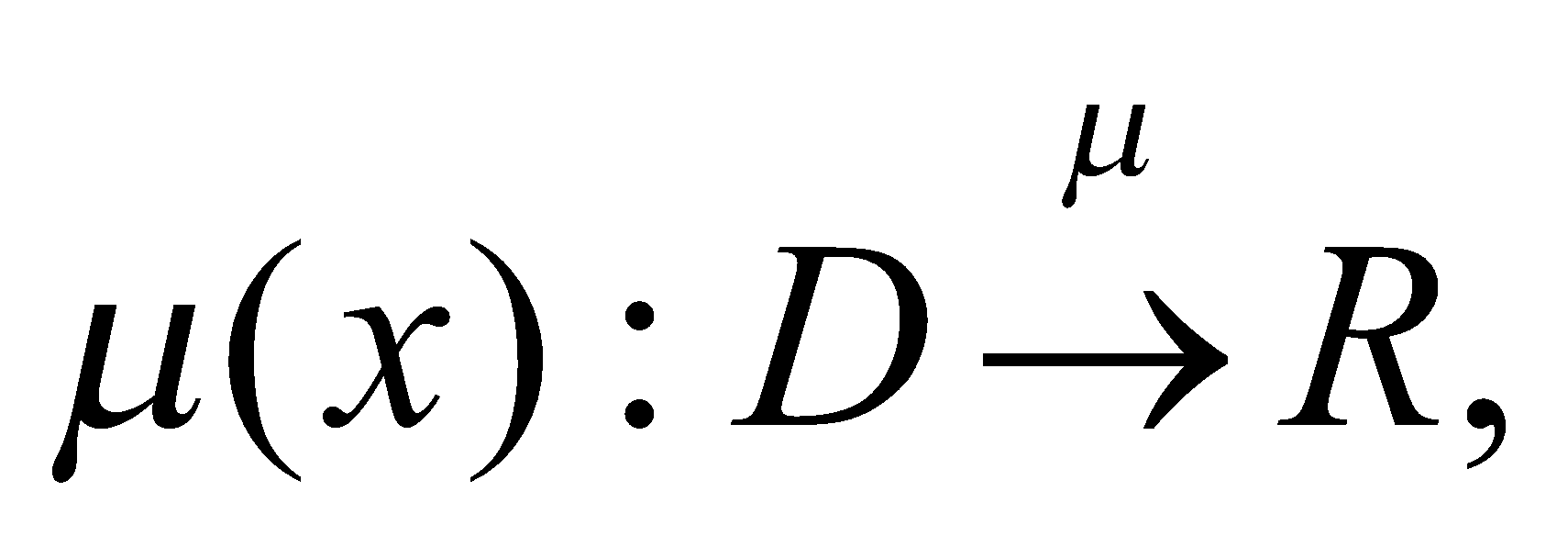
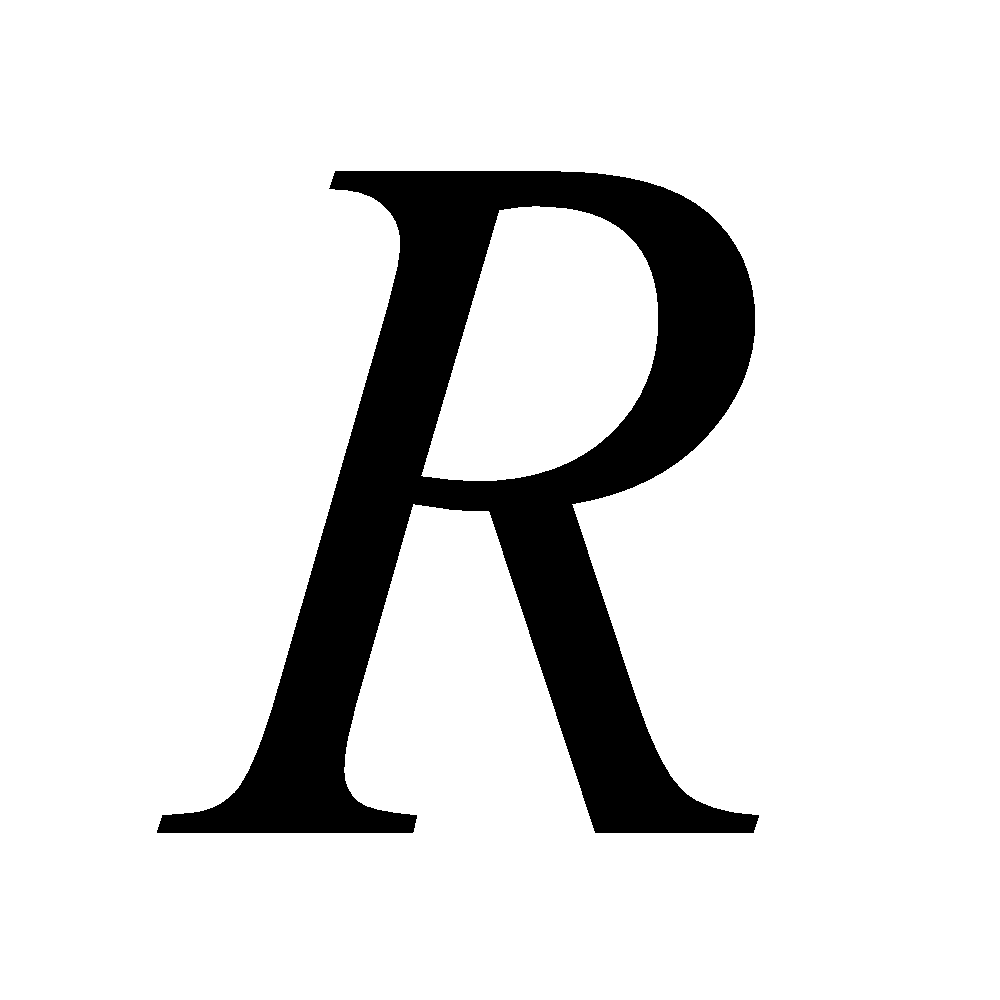
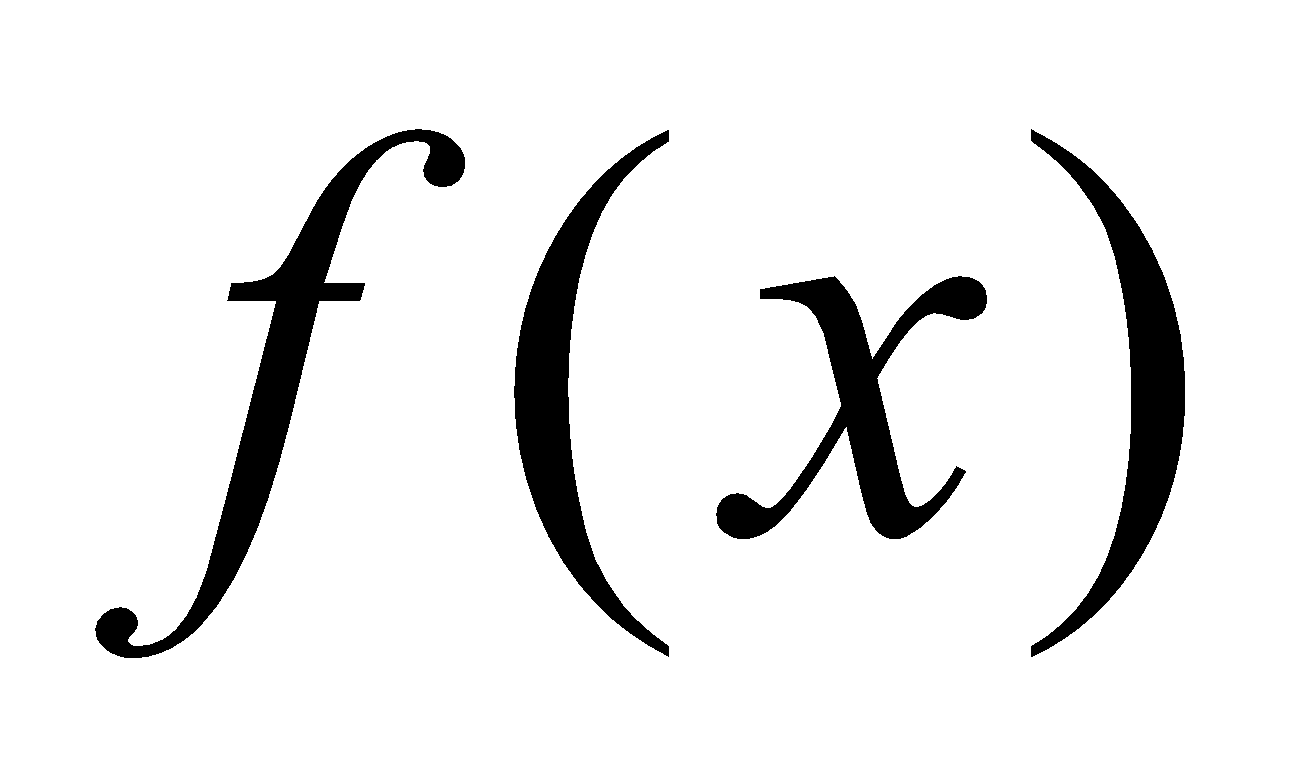
Возможных решений, называемых ***популяциями*,** где  - хромосомы с номером ,  - размер популяции,  - ген с номером - той популяции.

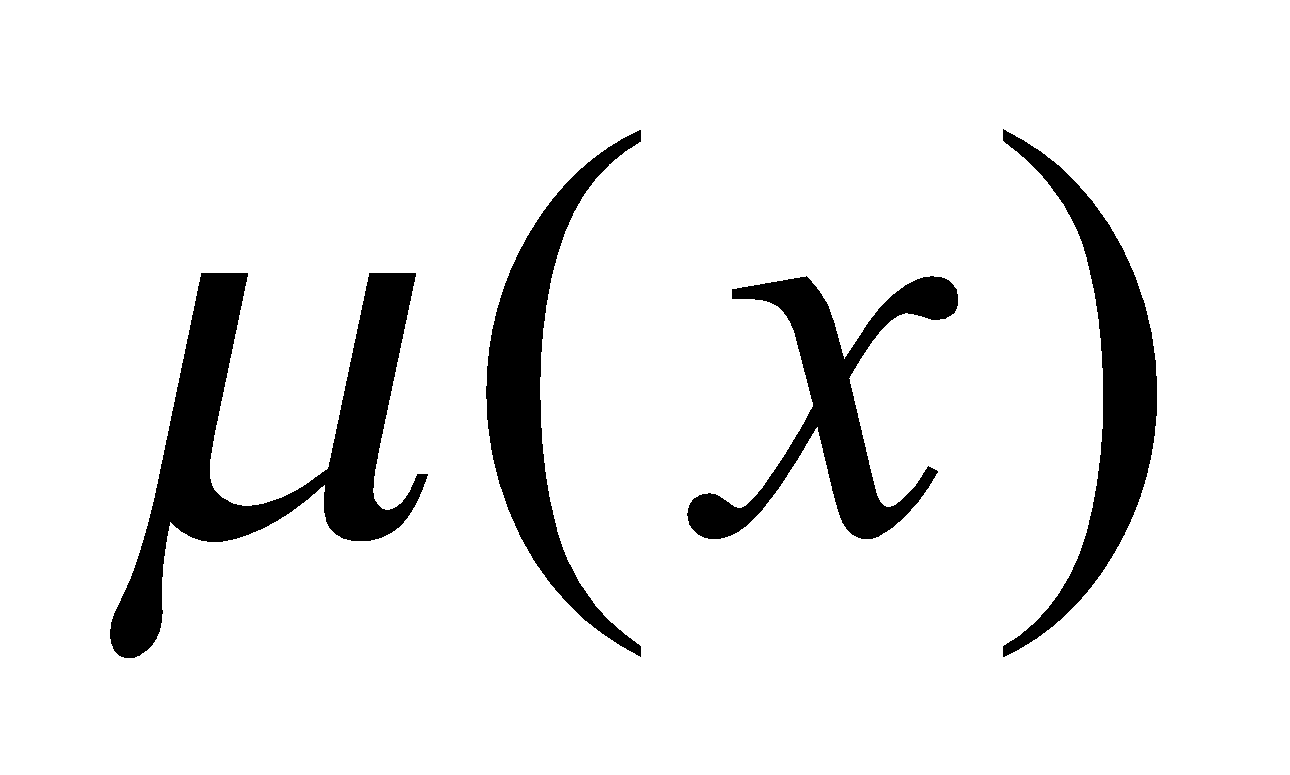
Затем осуществляется обратное преобразование:

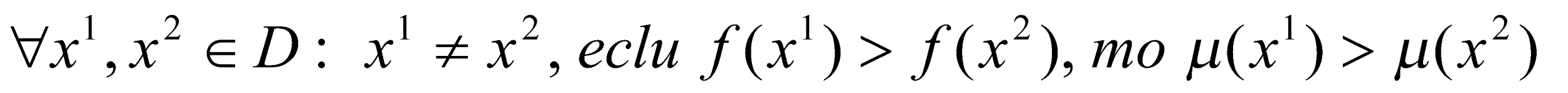
.

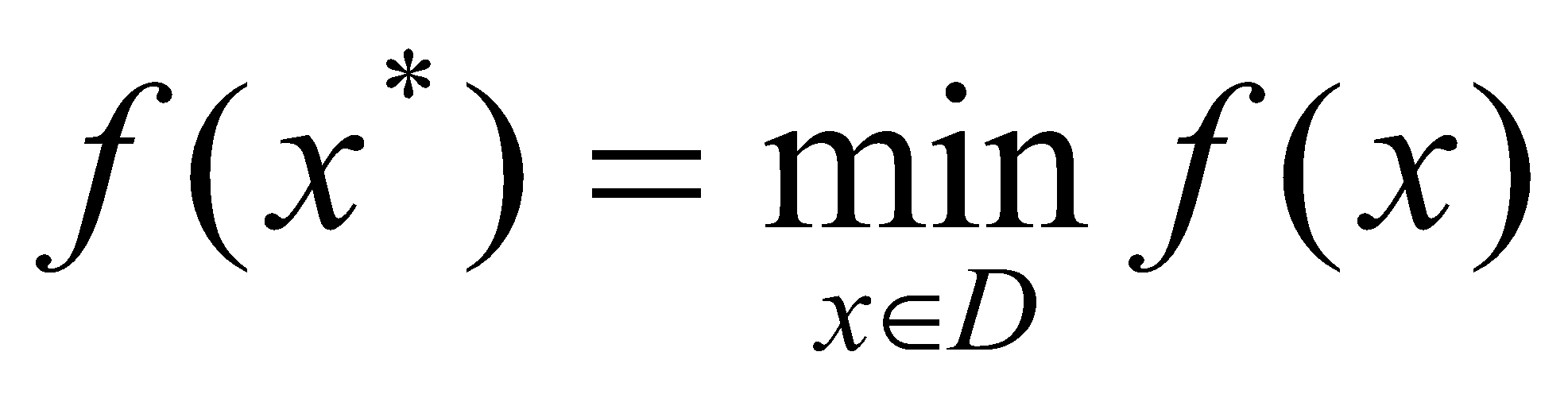
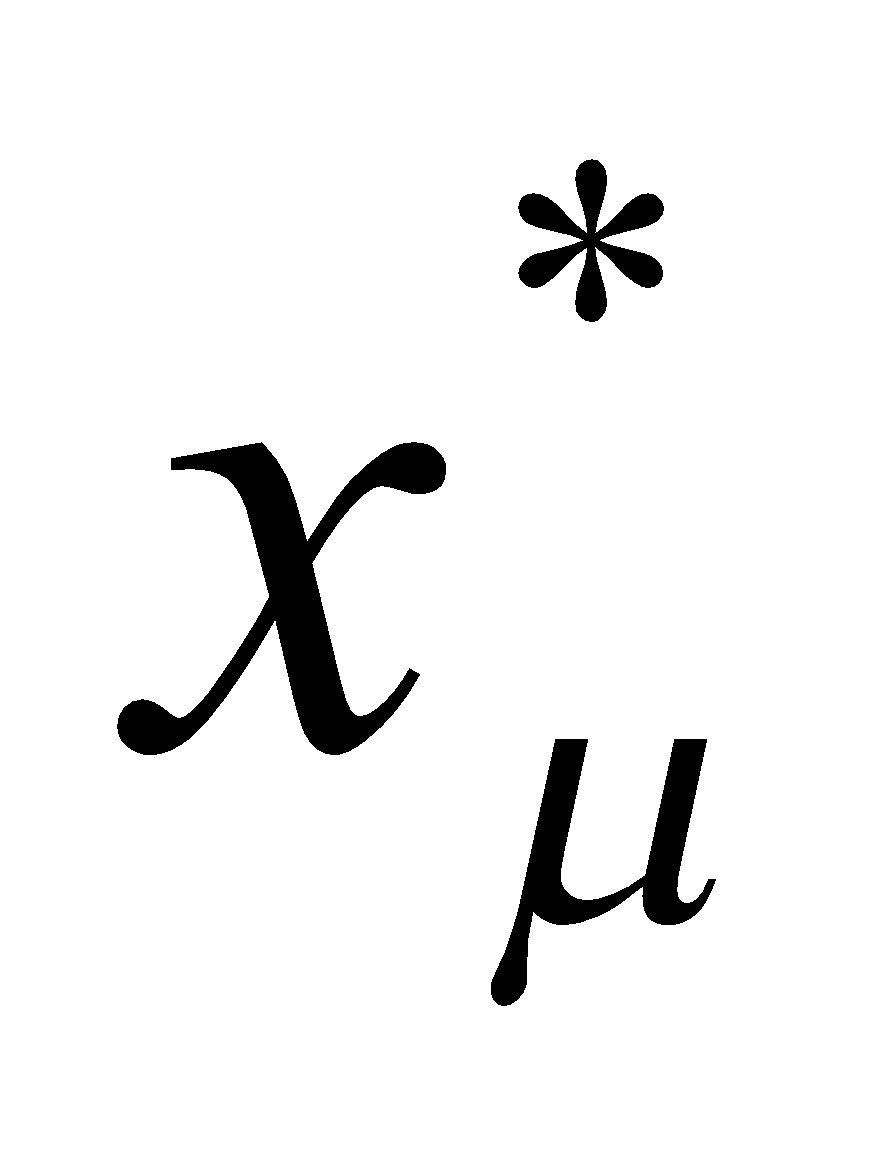
Различают два способа кодирования:

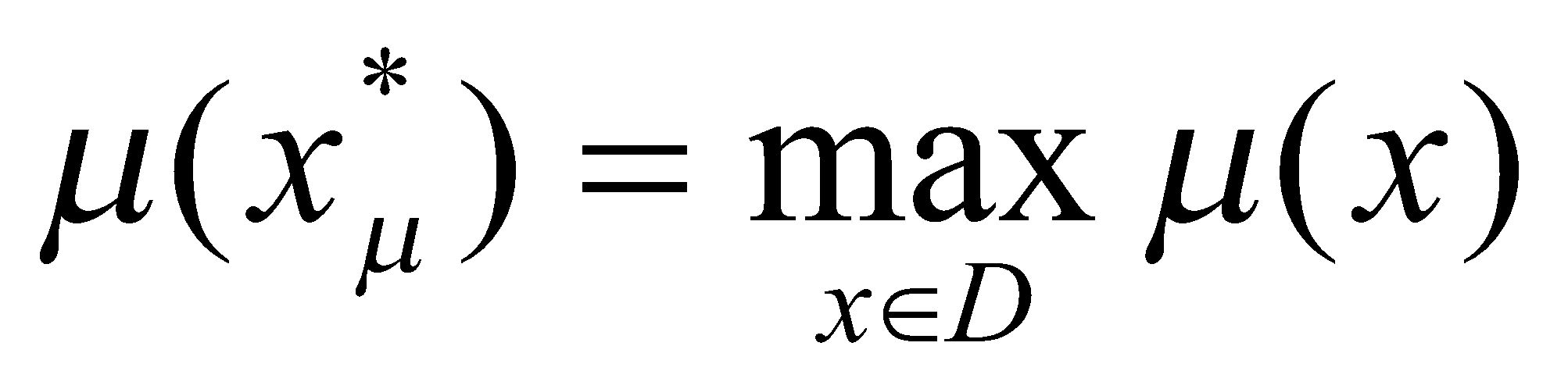
1. Бинарное кодирование.
2. Вещественное кодирование.

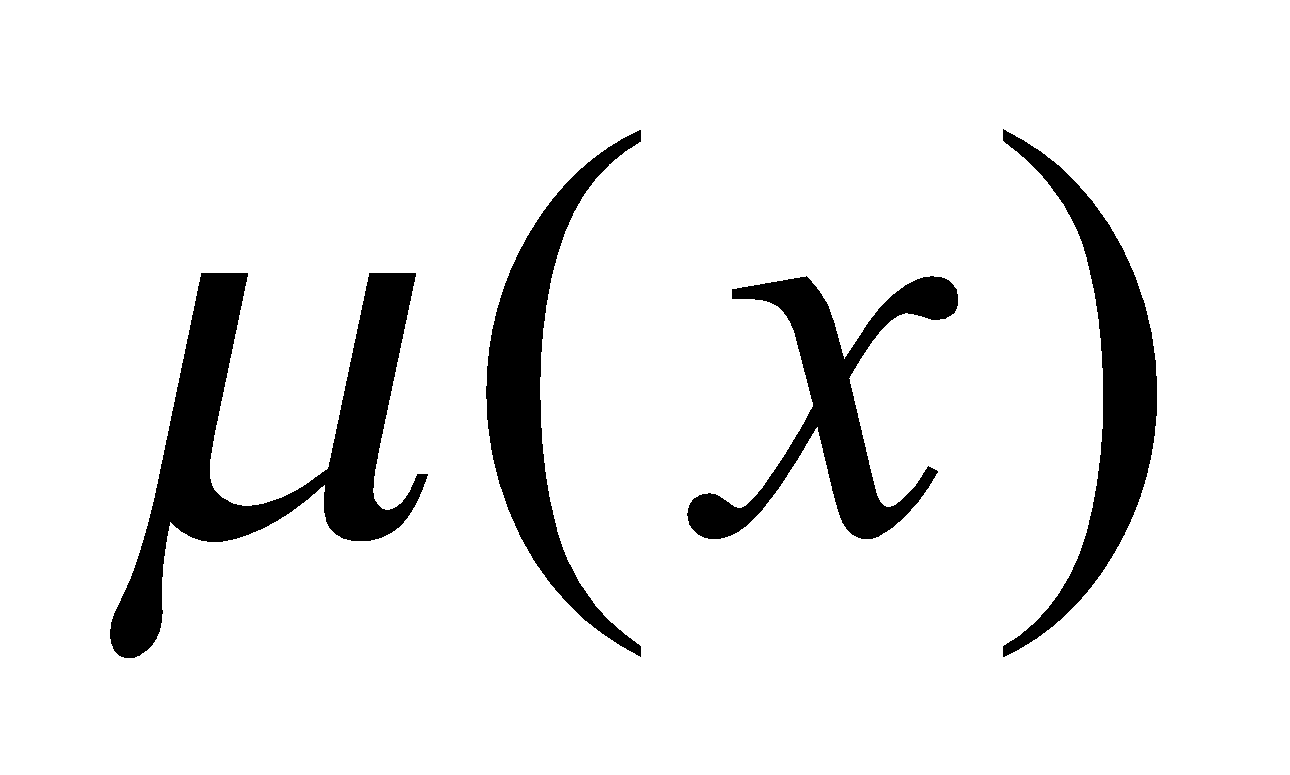
Будем использовать второй вариант кодирования. В этом случае целевая функция может использоваться непосредственно как функция фитнеса. Тогда в качестве функции фитнеса  получается как преобразование целевой функции, т.е. *функция фитнеса*  где  - множество вещественных чисел, аналогична целевой функции .

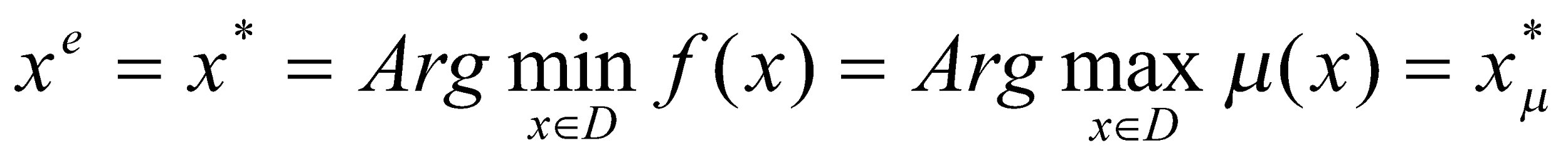
Функцией  может быть любая функция, удовлетворяющая следующему условию:

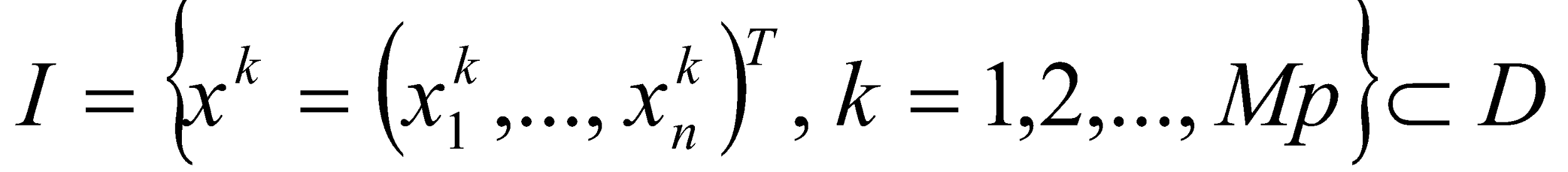
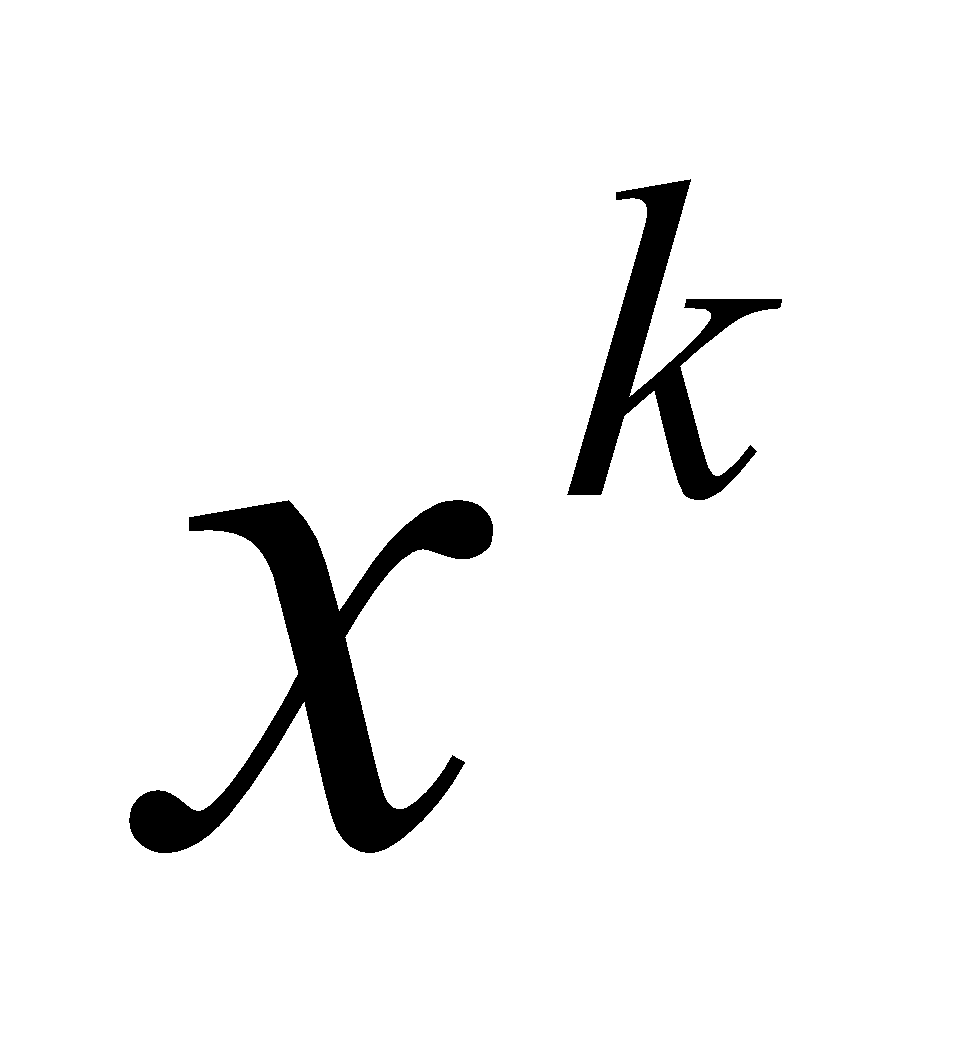
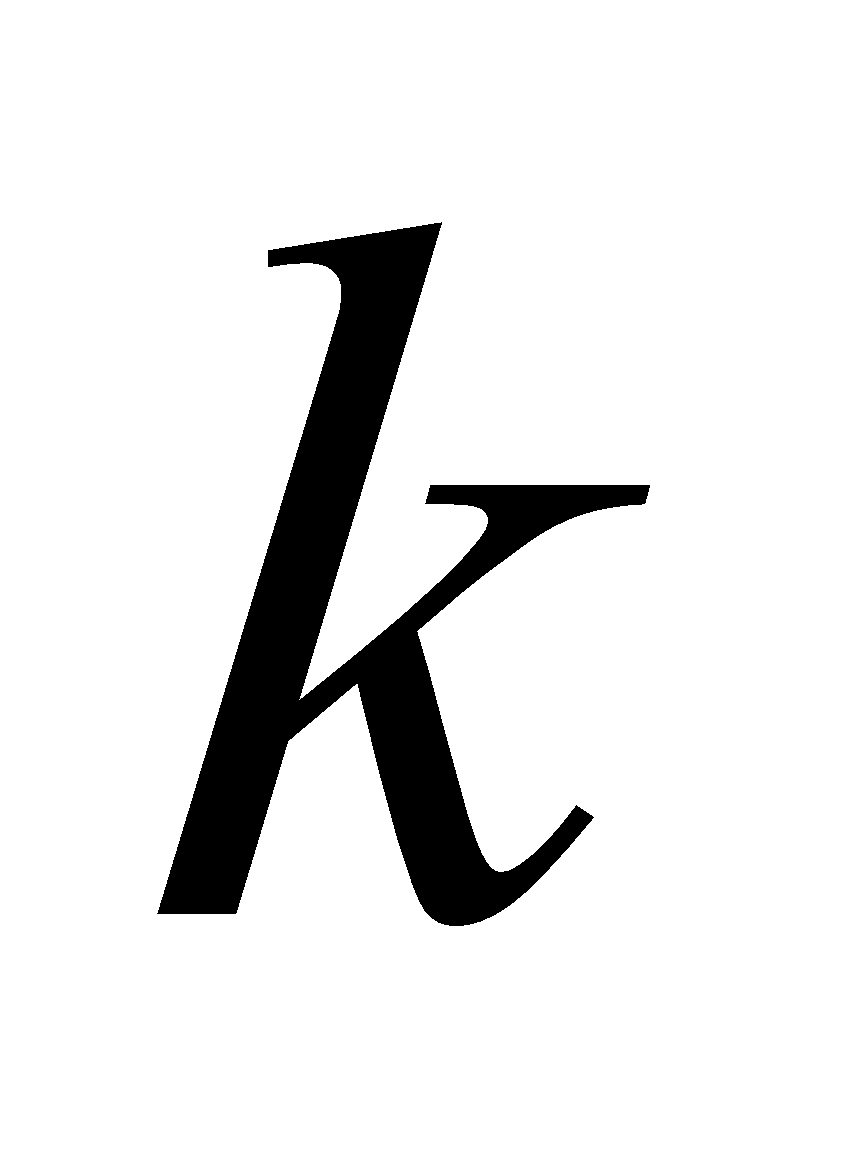
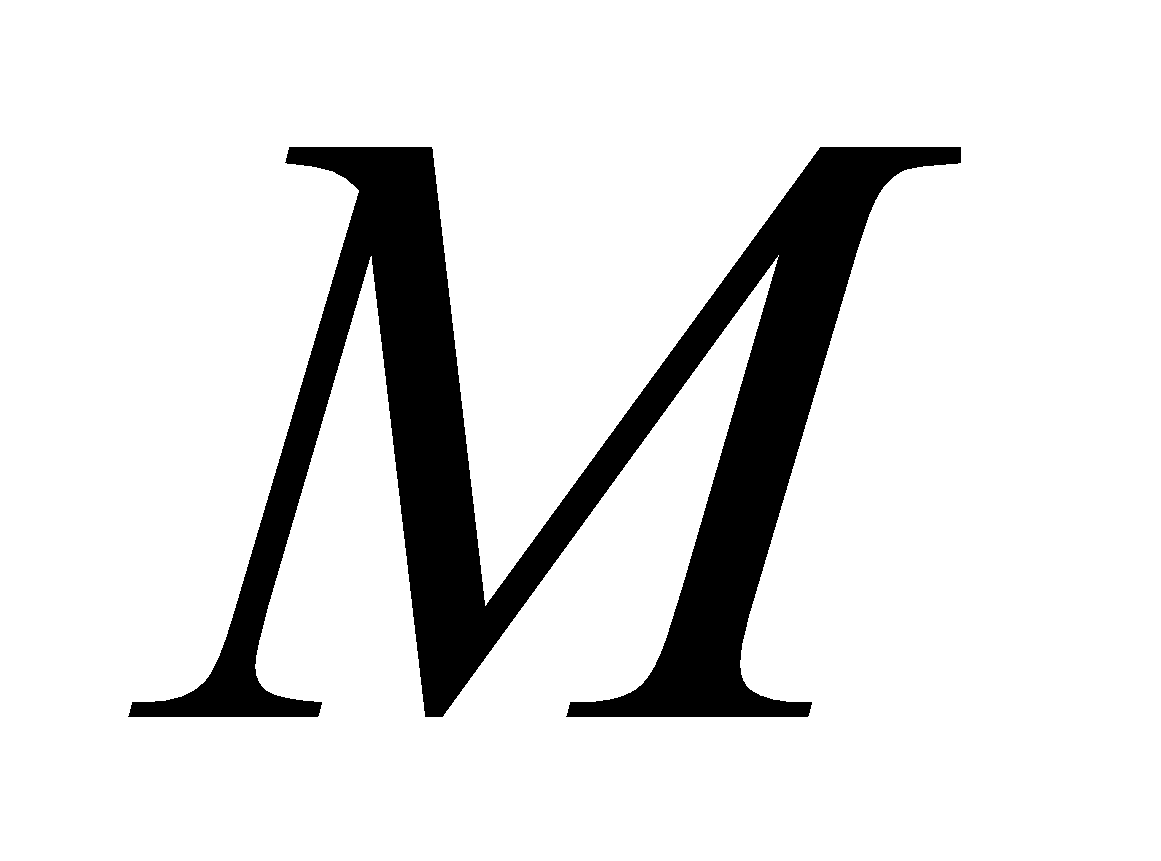
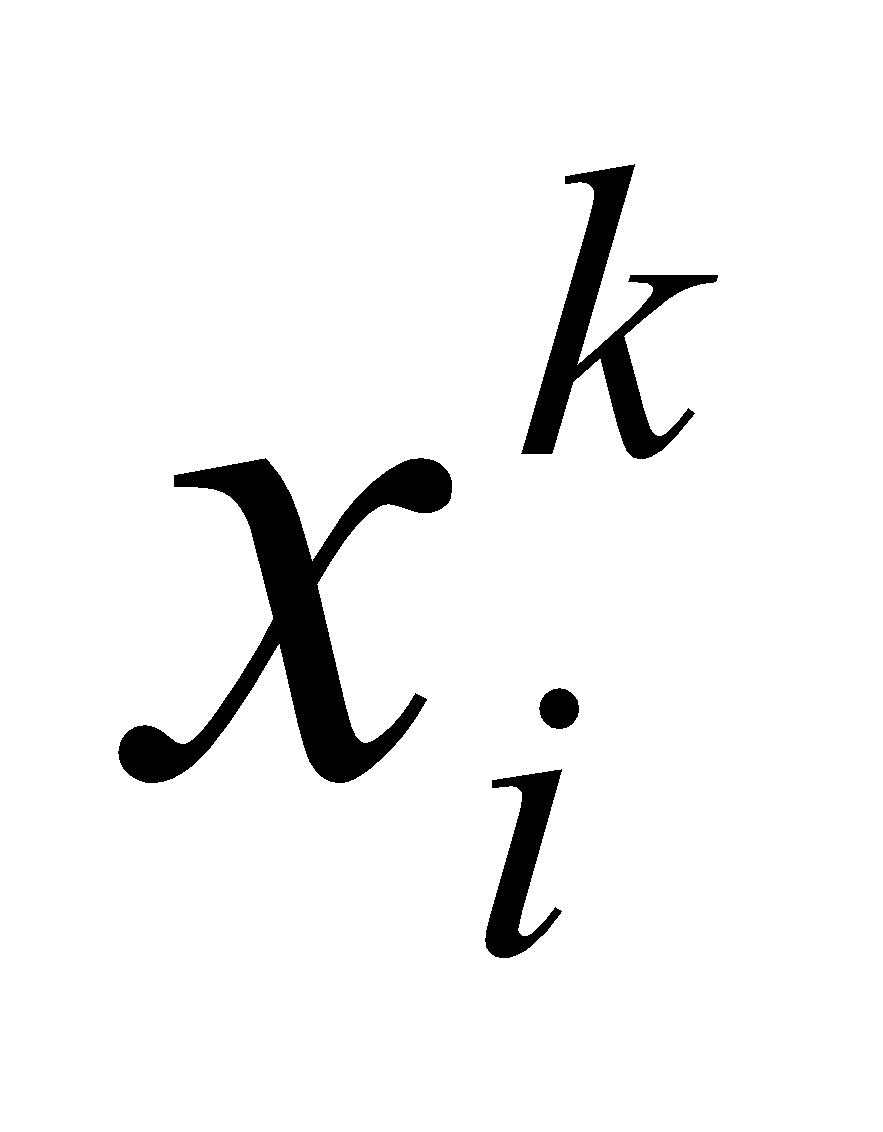
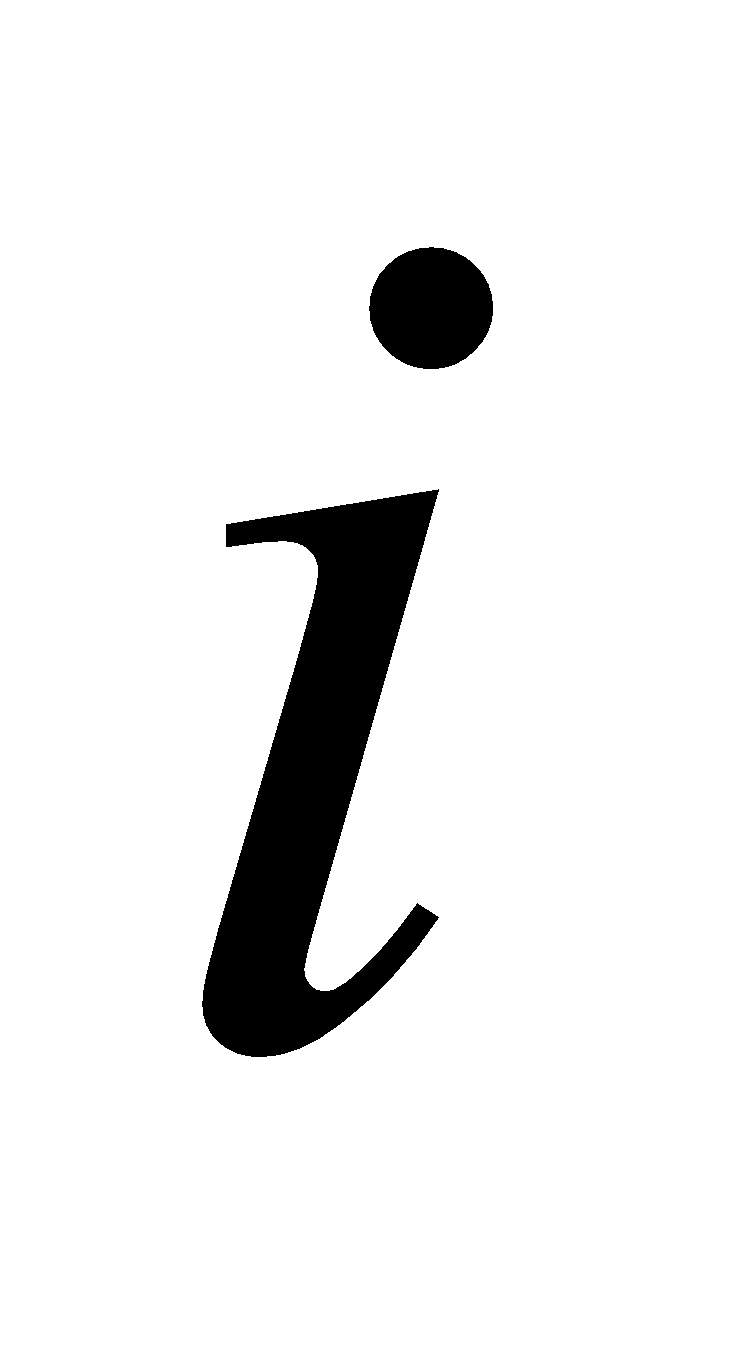
.

Решение исходной оптимизационной задачи  сводится к поиску решения  другой оптимизационной задачи:

. (2)

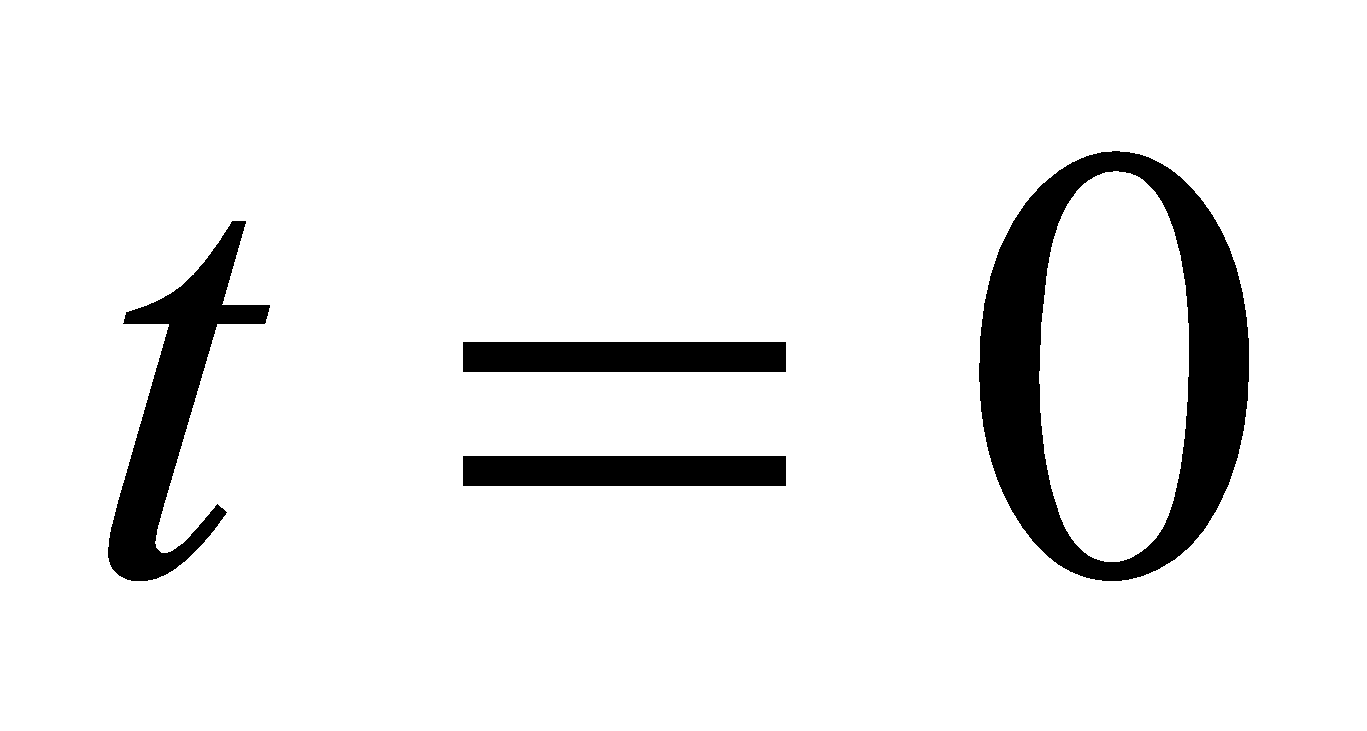
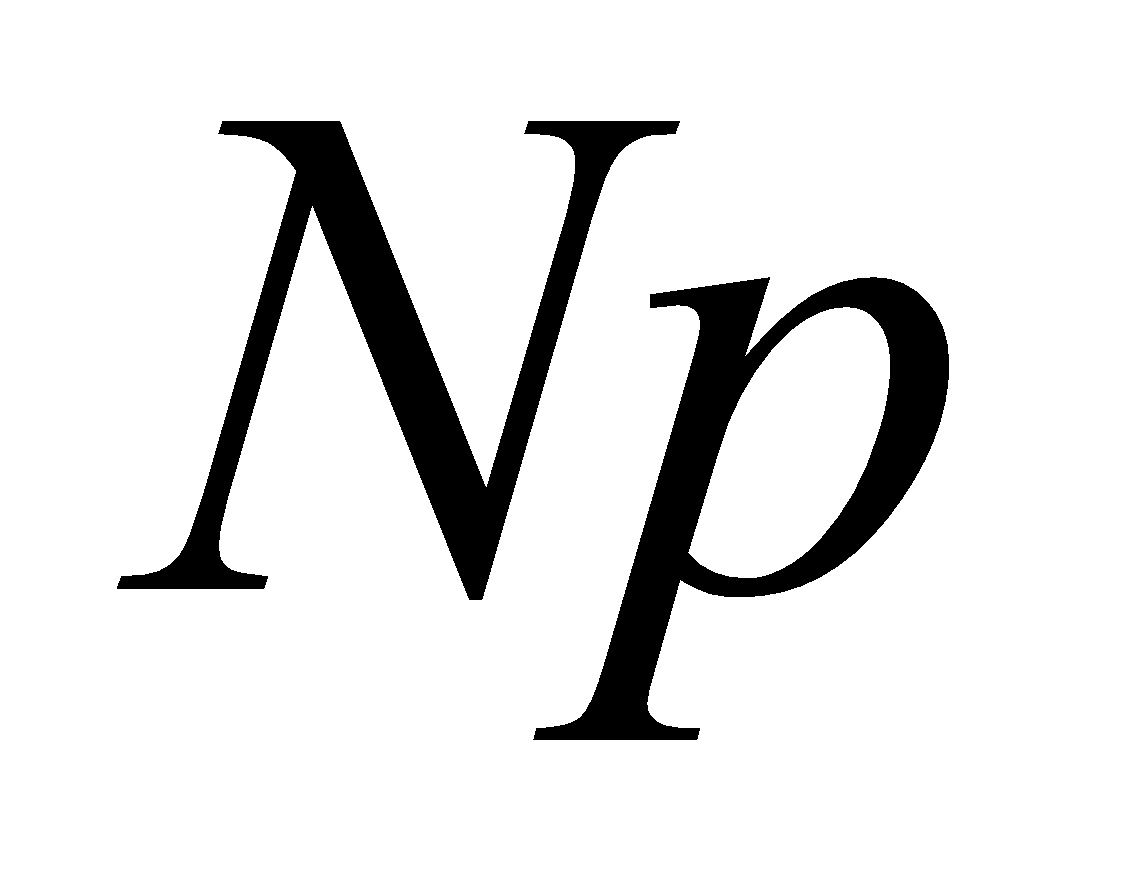
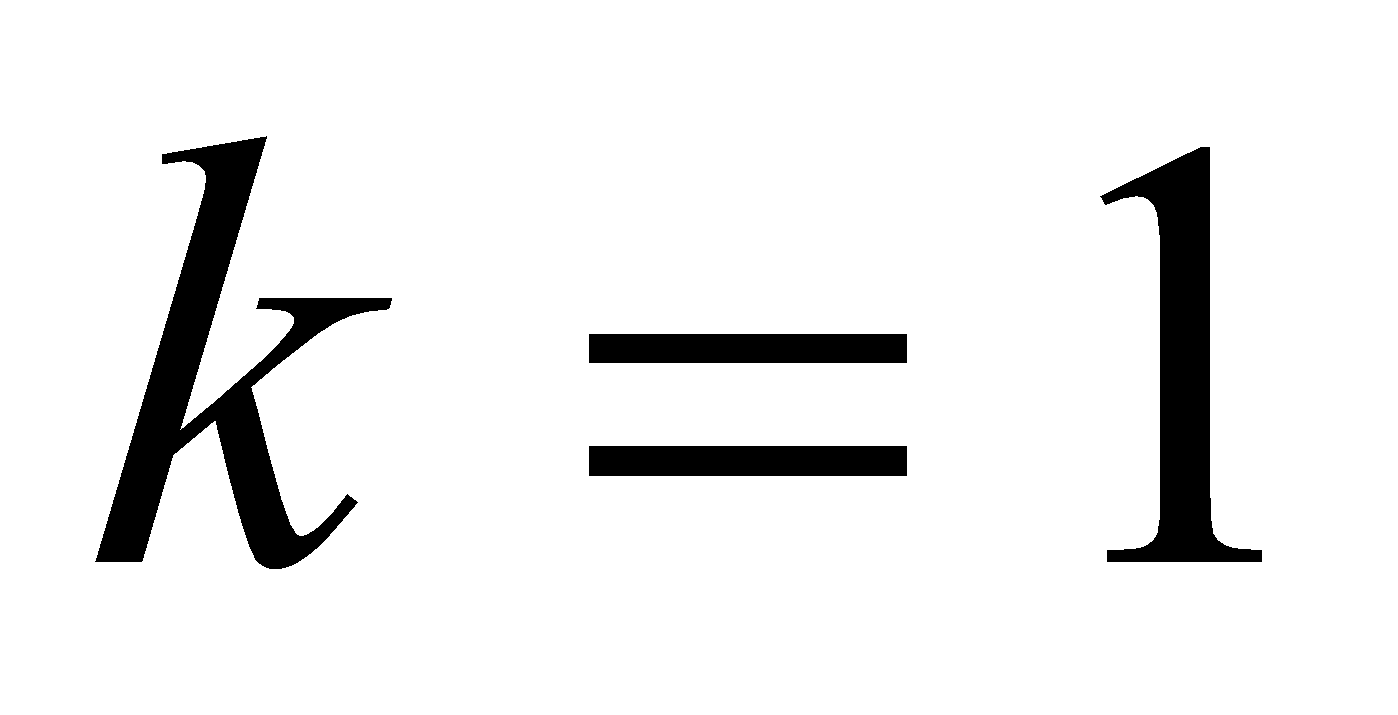
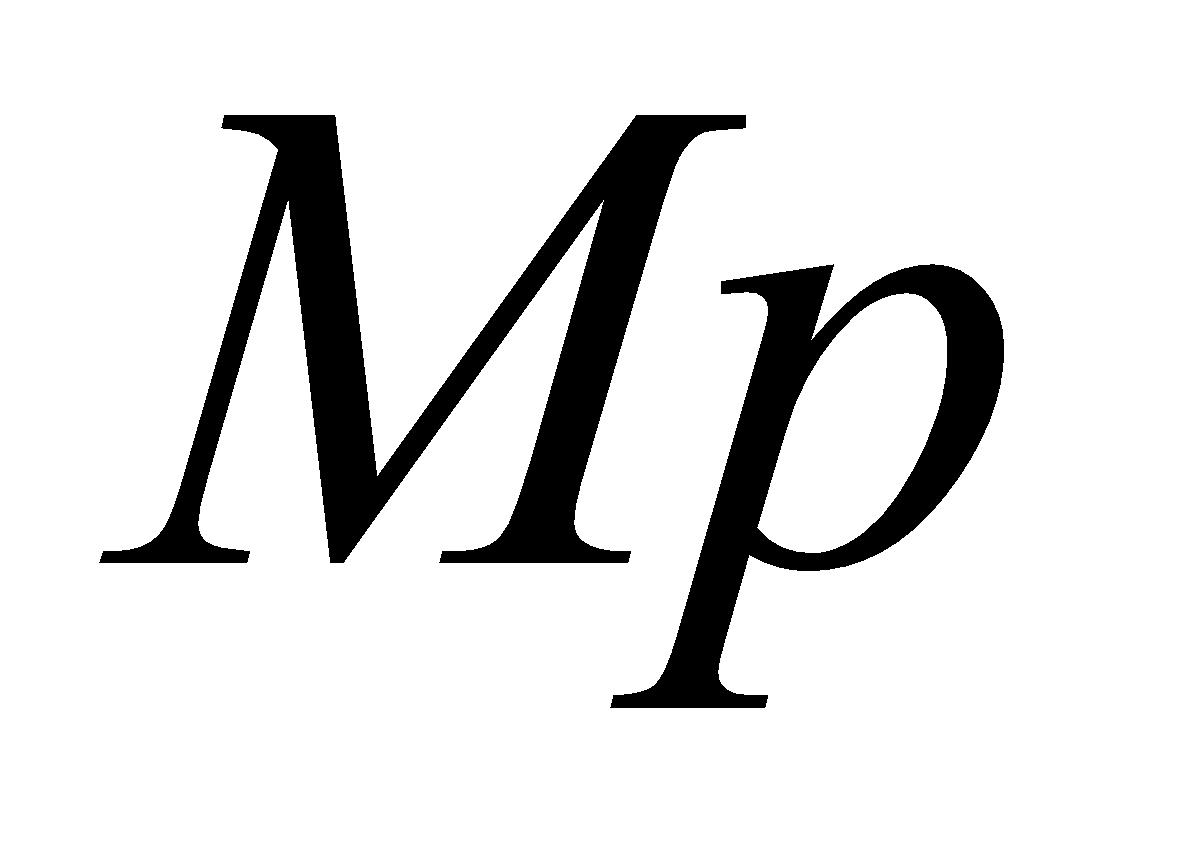
В силу выбора функции , решение задач (1) и (2) (хромосома) совпадают:

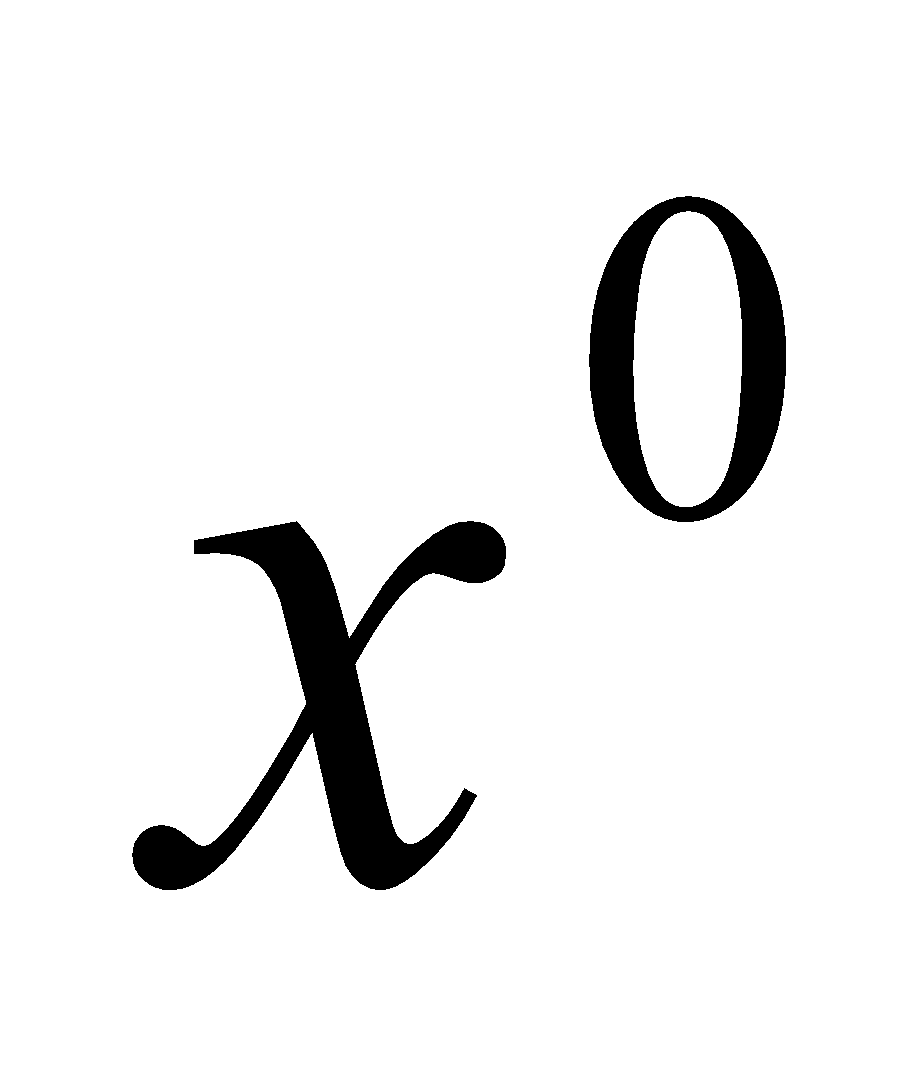
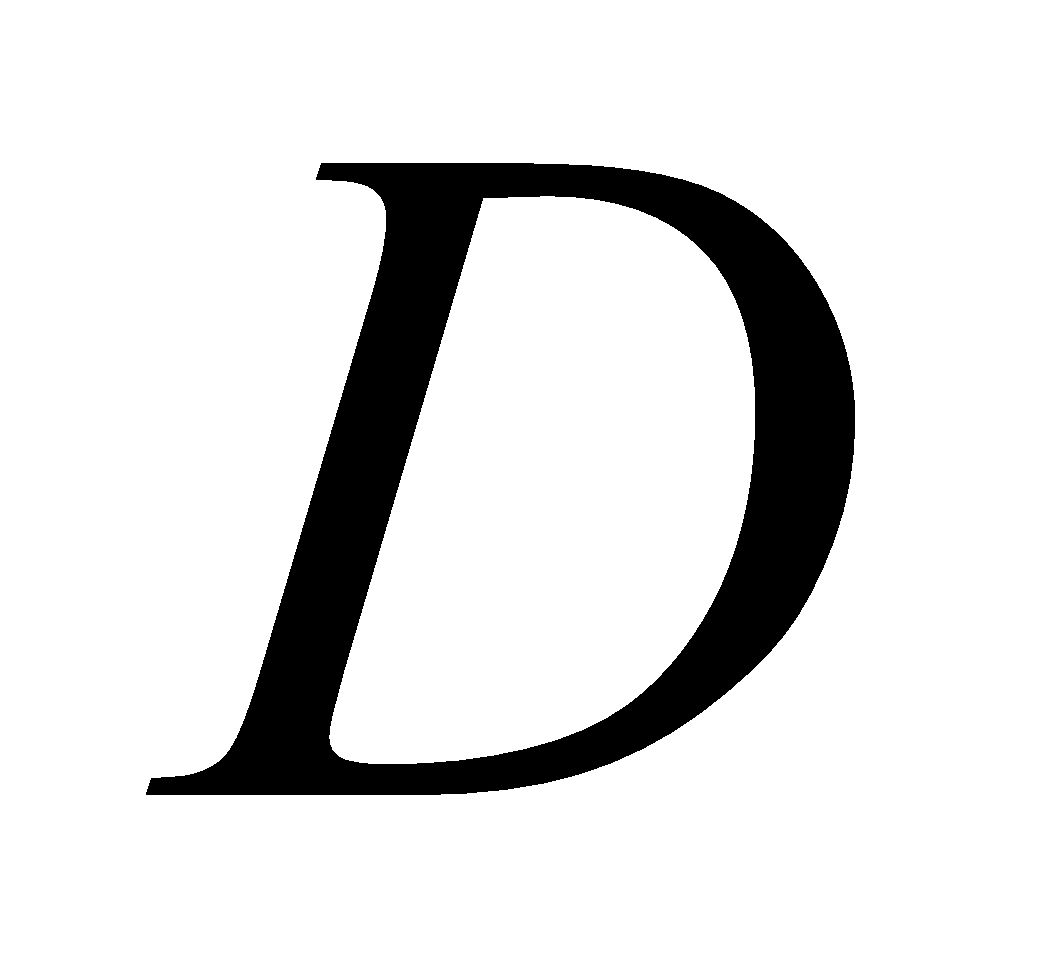
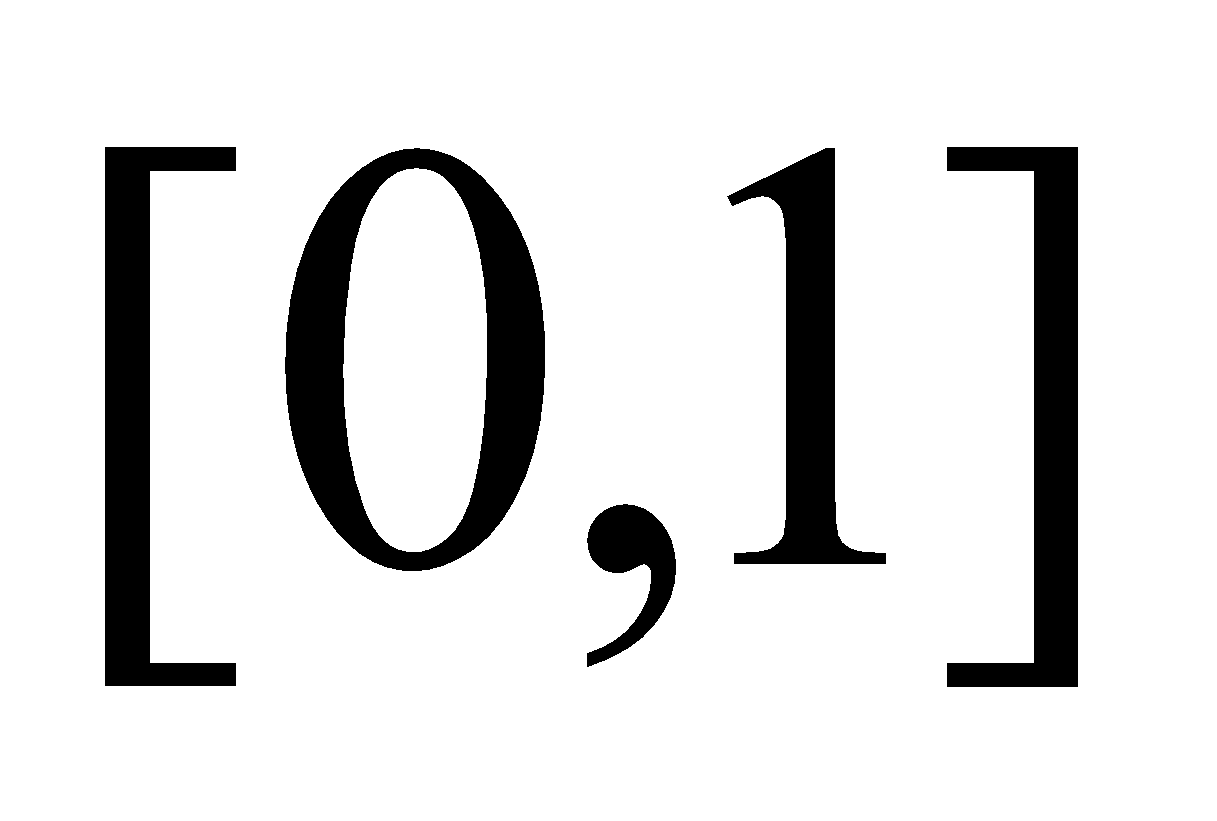
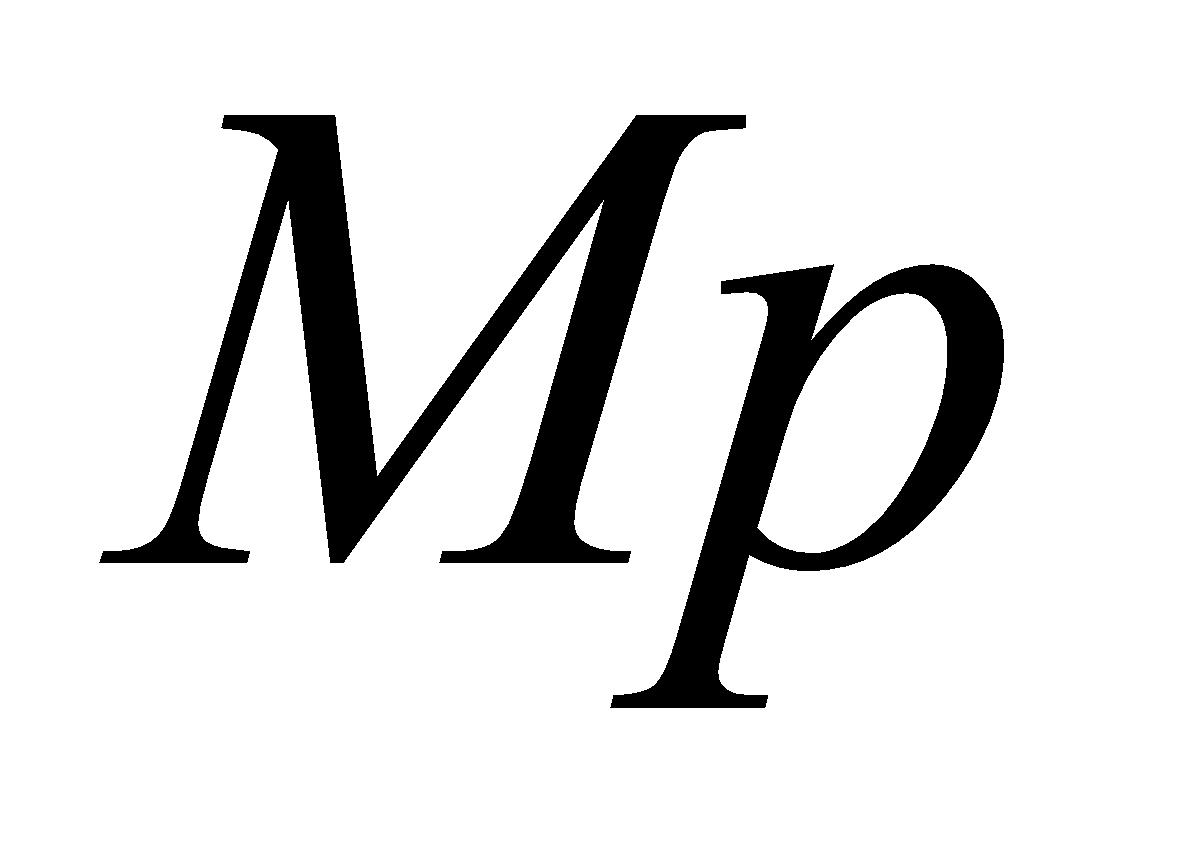
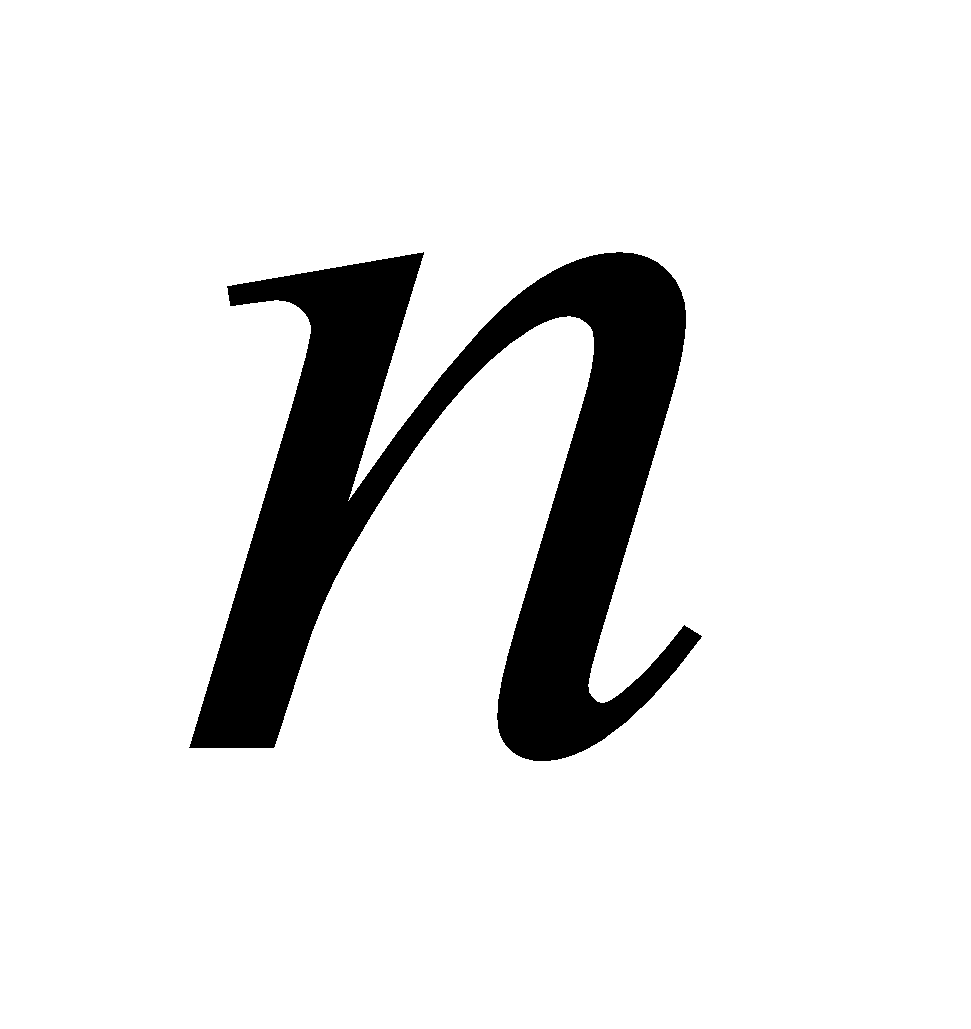
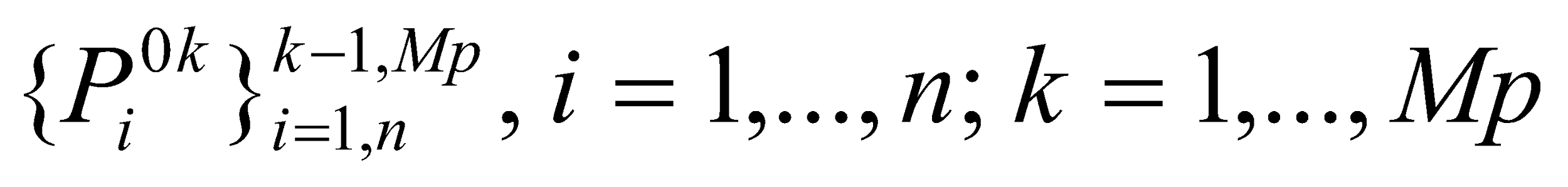
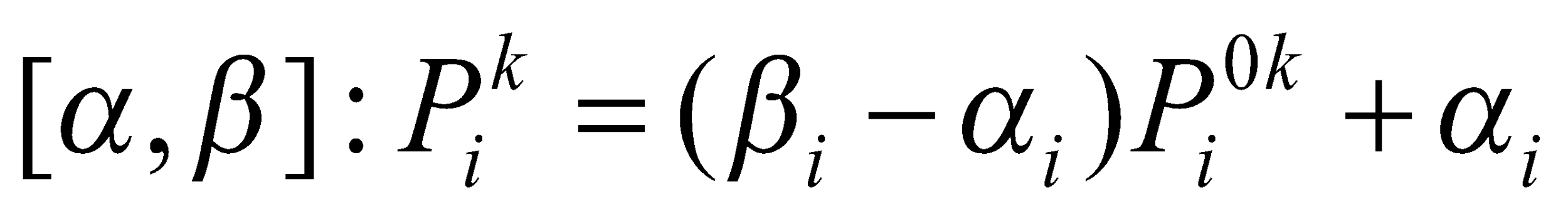
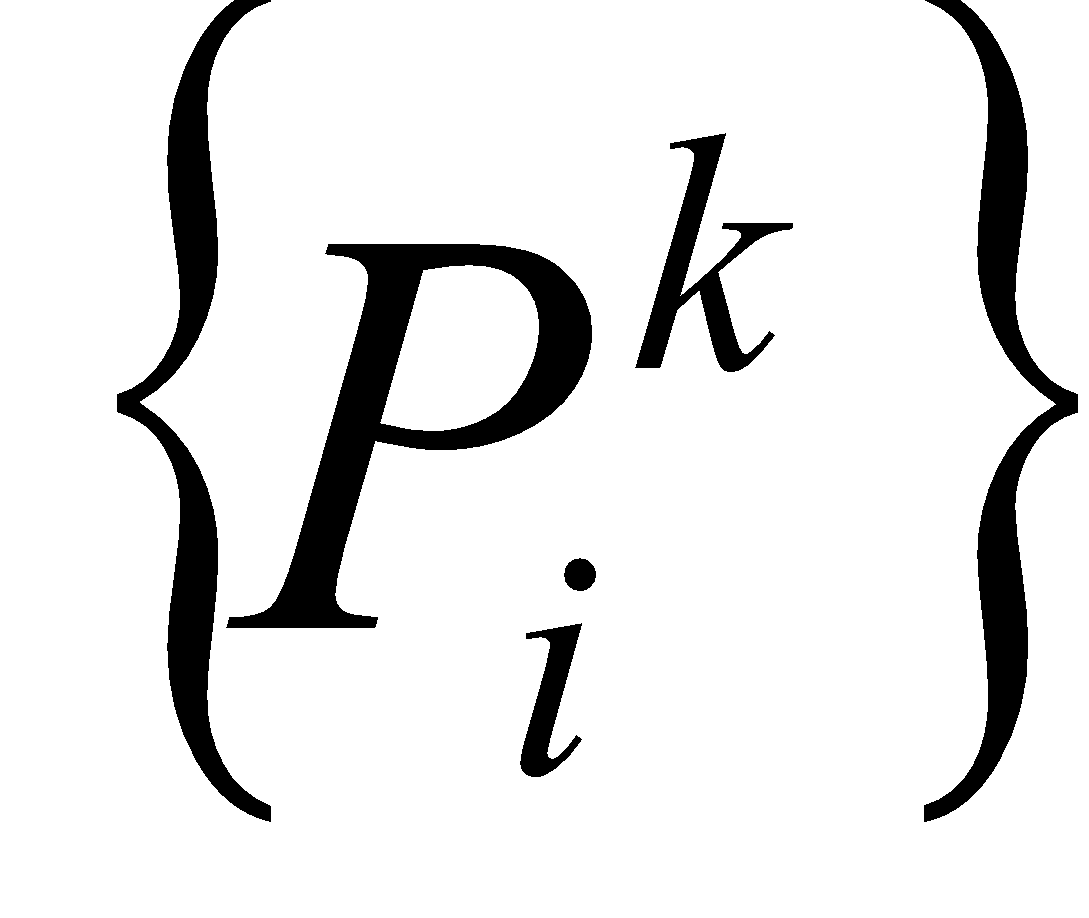
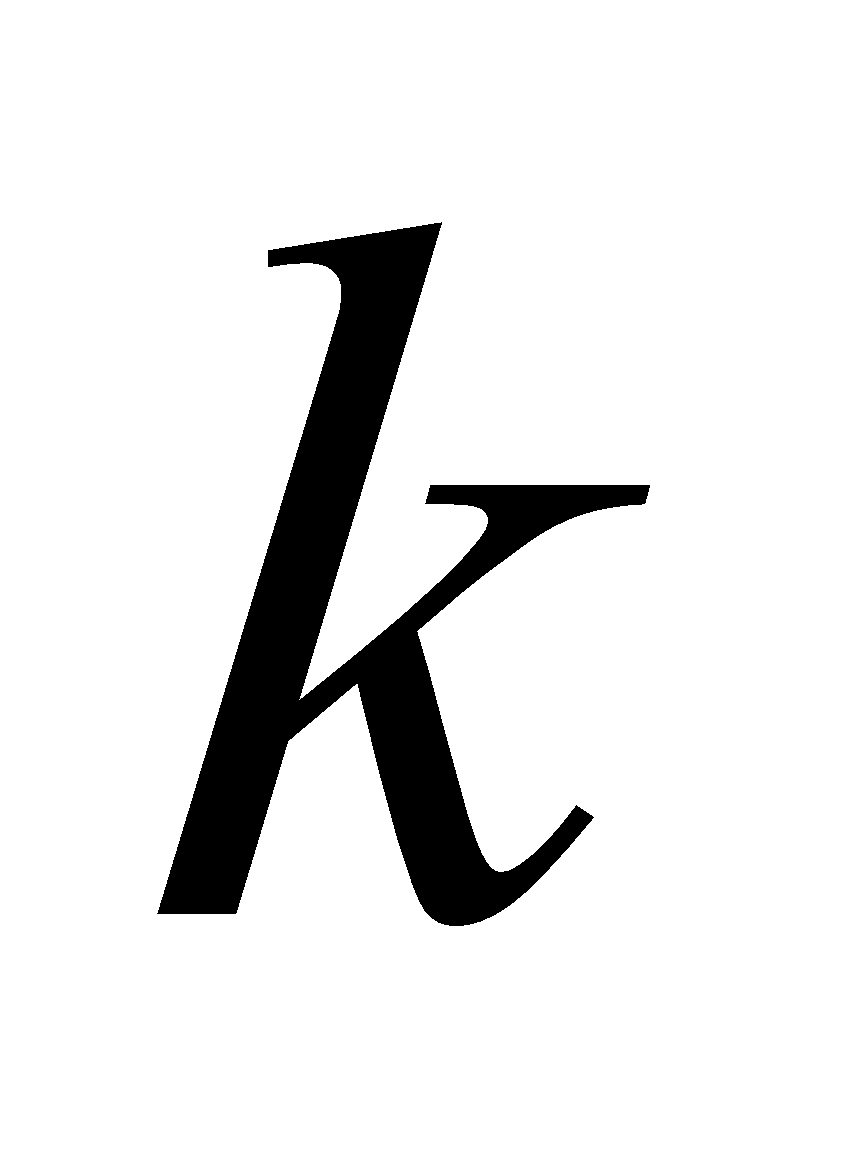
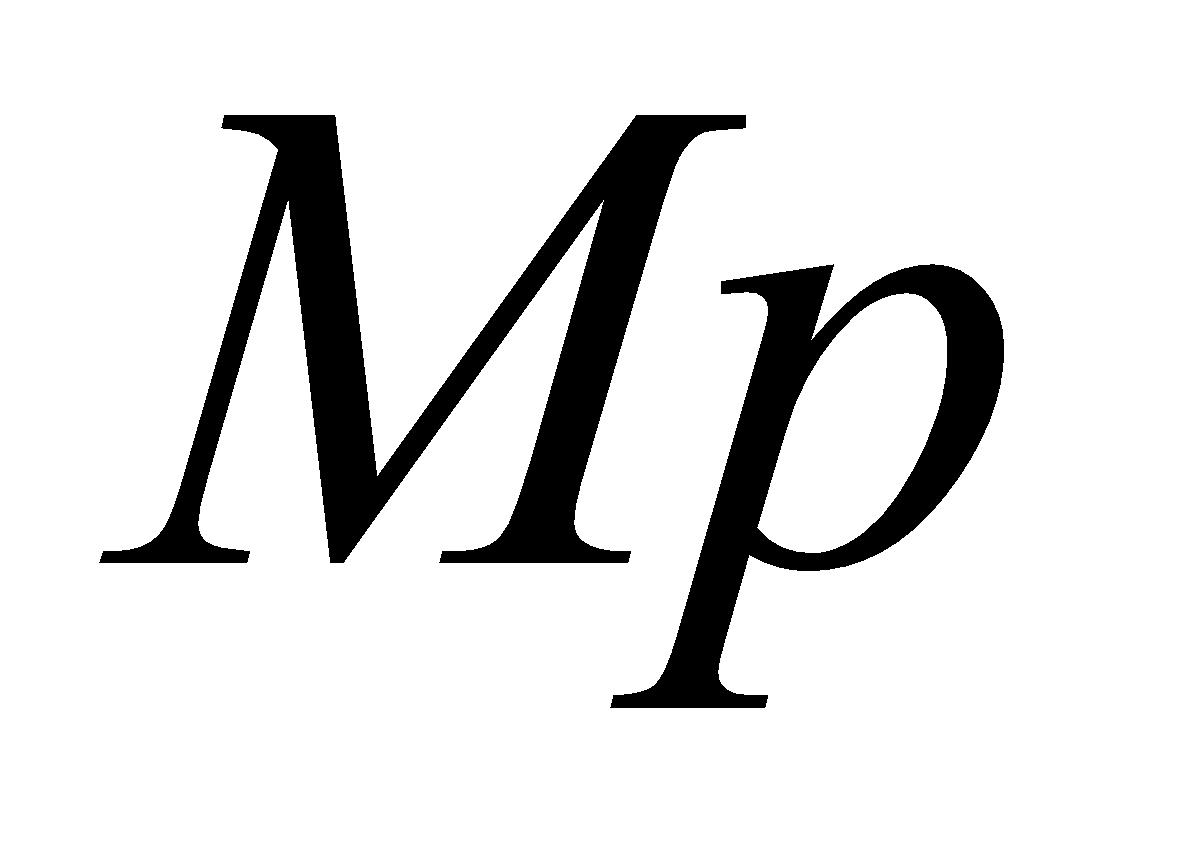
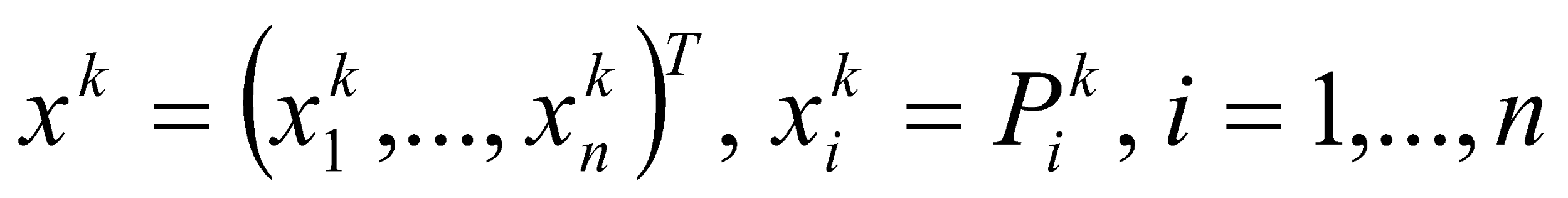
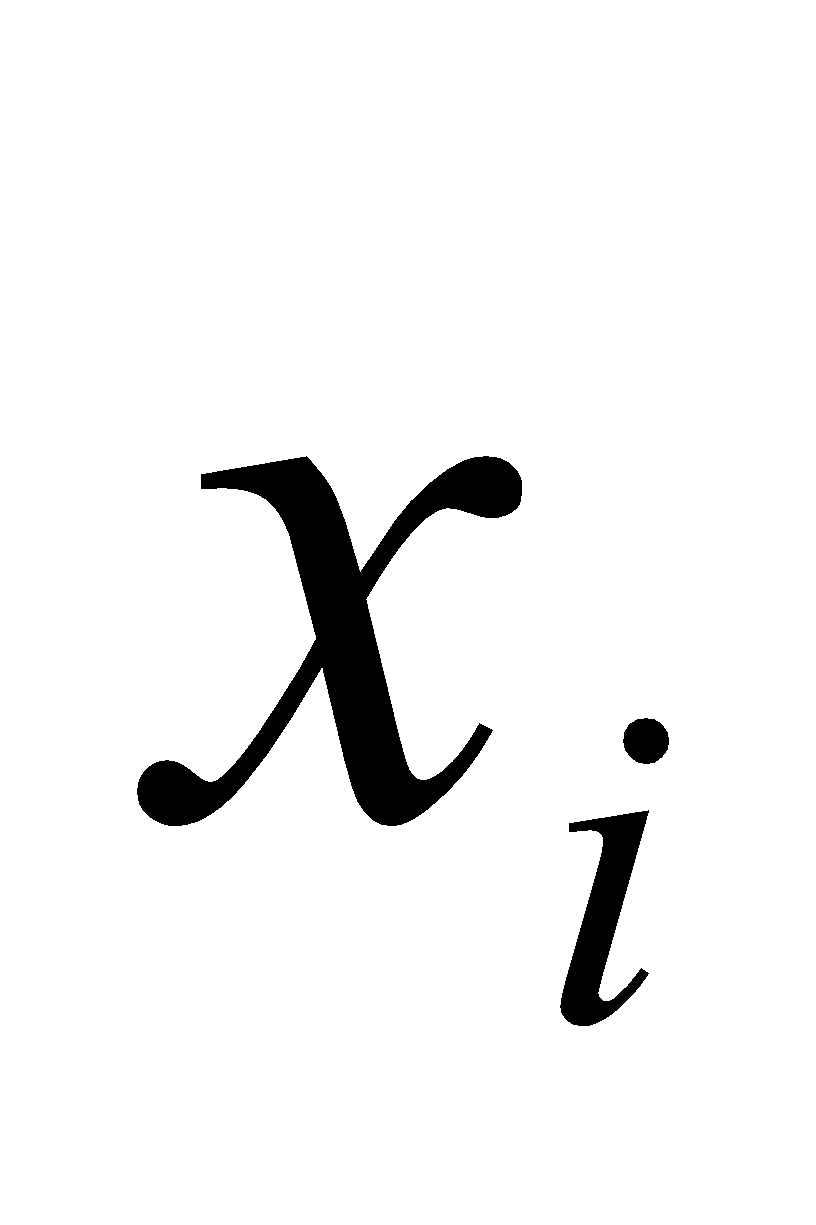
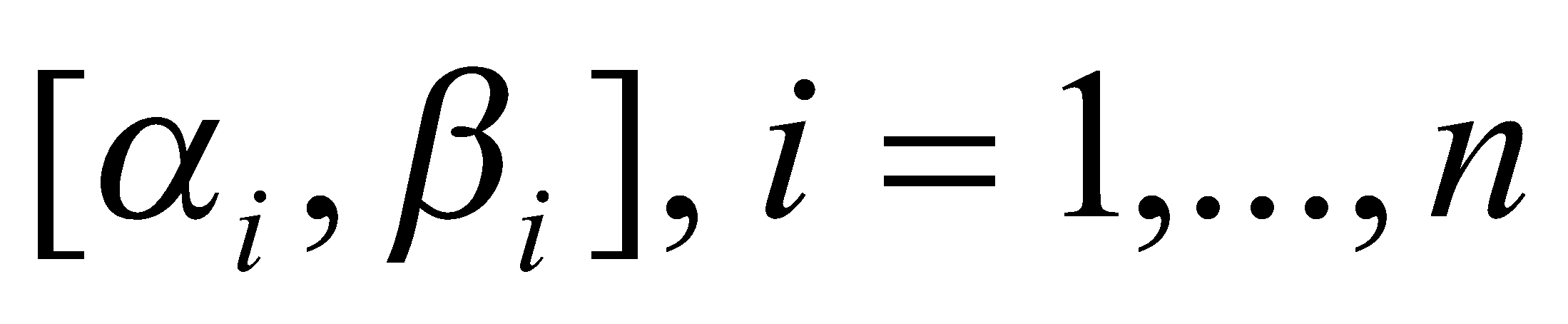
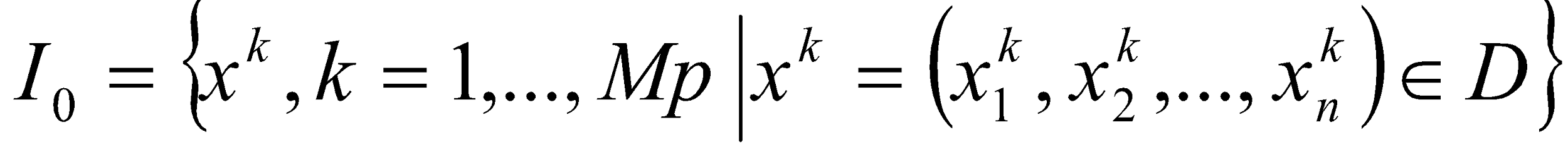
. (3)

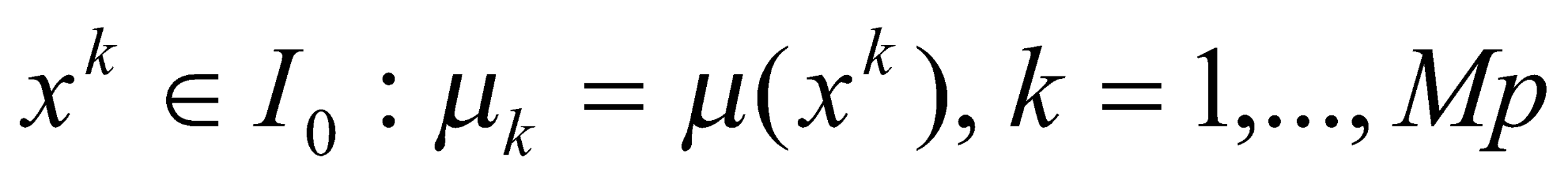
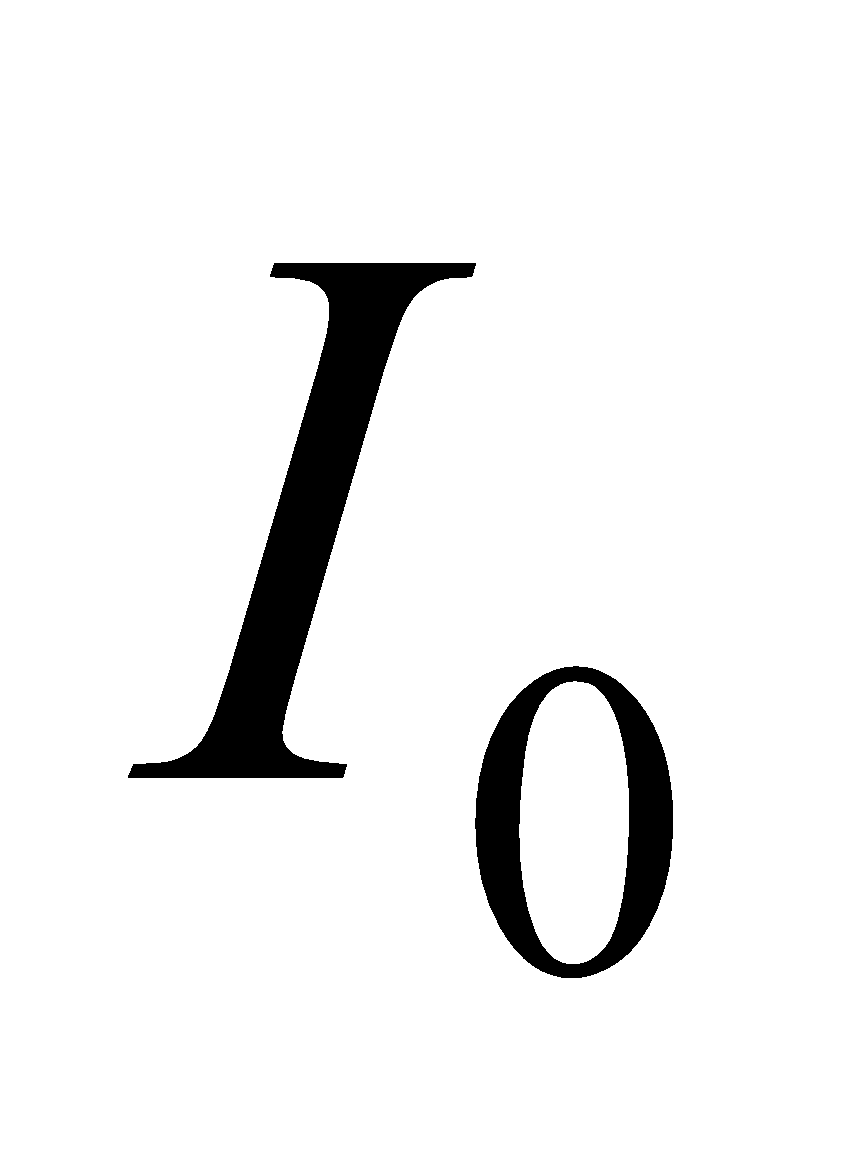
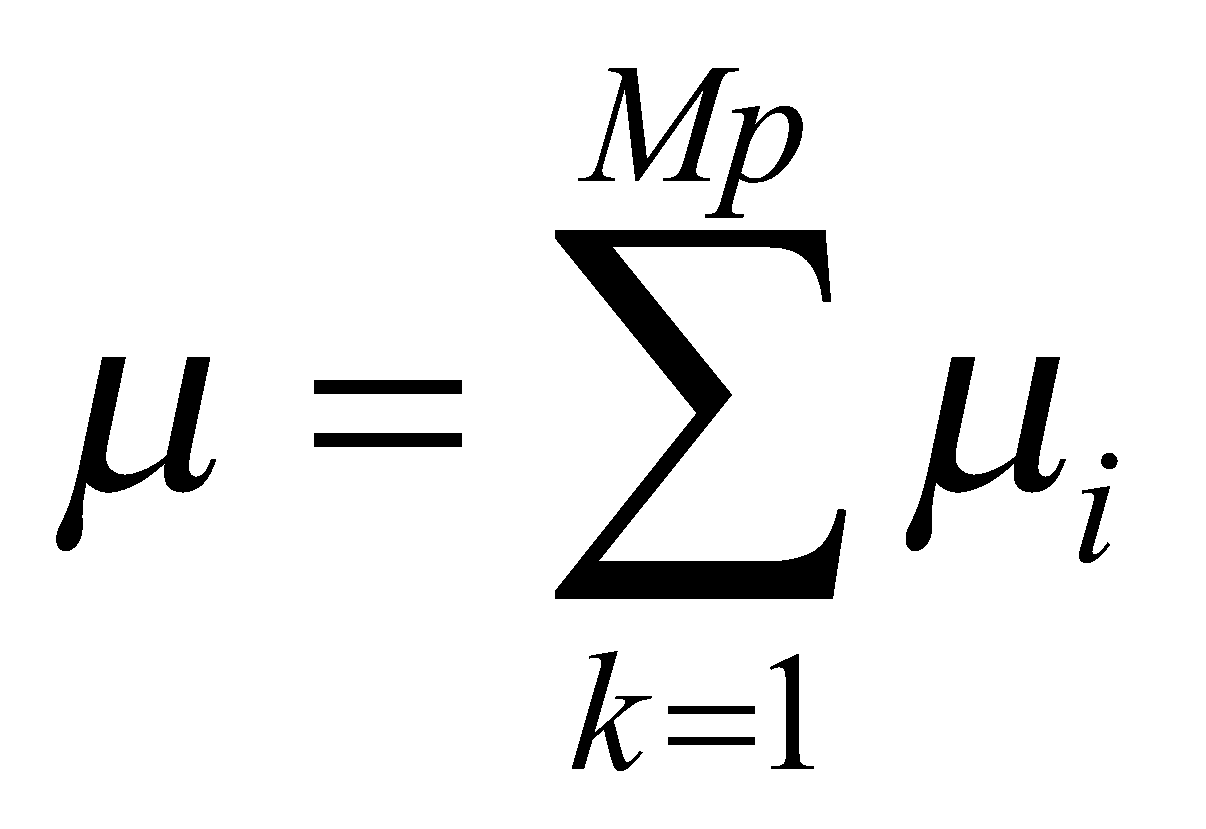
При решении задачи (2) используются конечные наборы  возможных решений, называемых *популяциями*, где  - хромосома с номером ,  - размер популяции,  - ген с номером .

* 1. **Генетический алгоритм.**

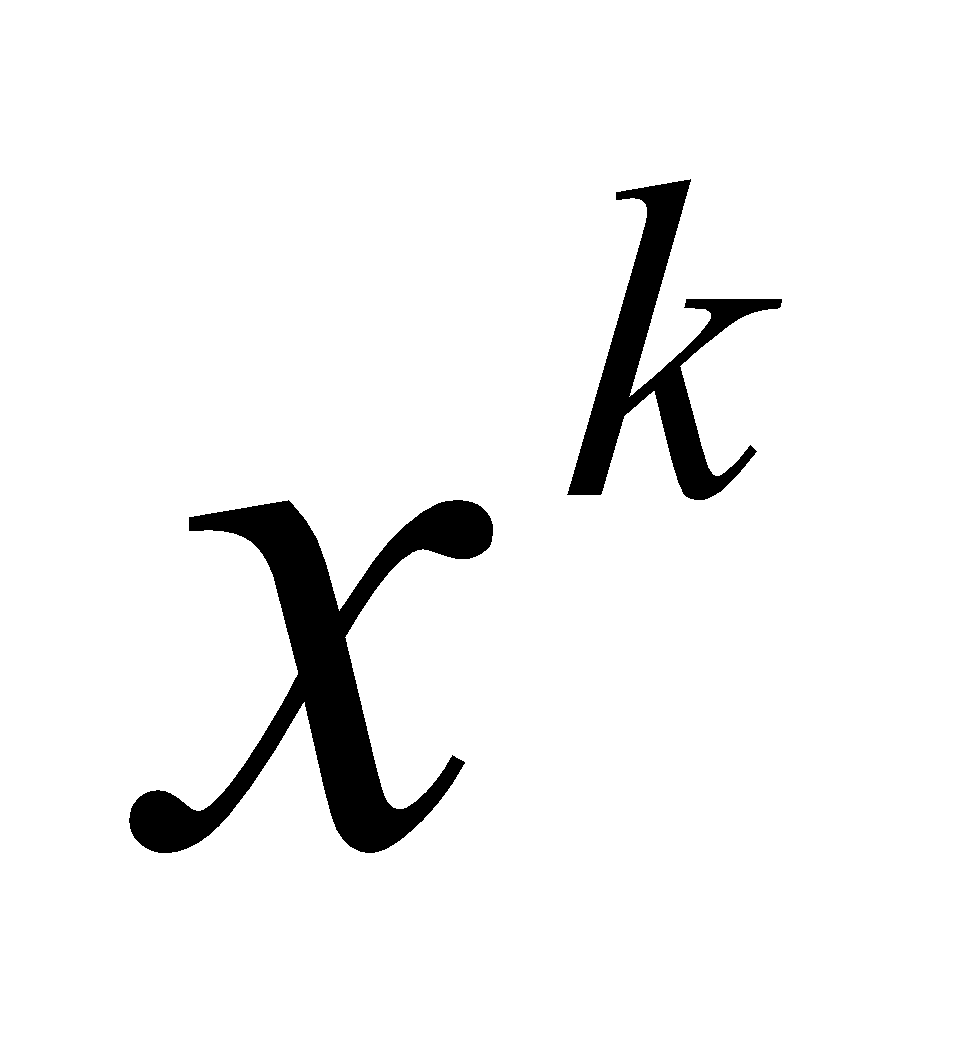
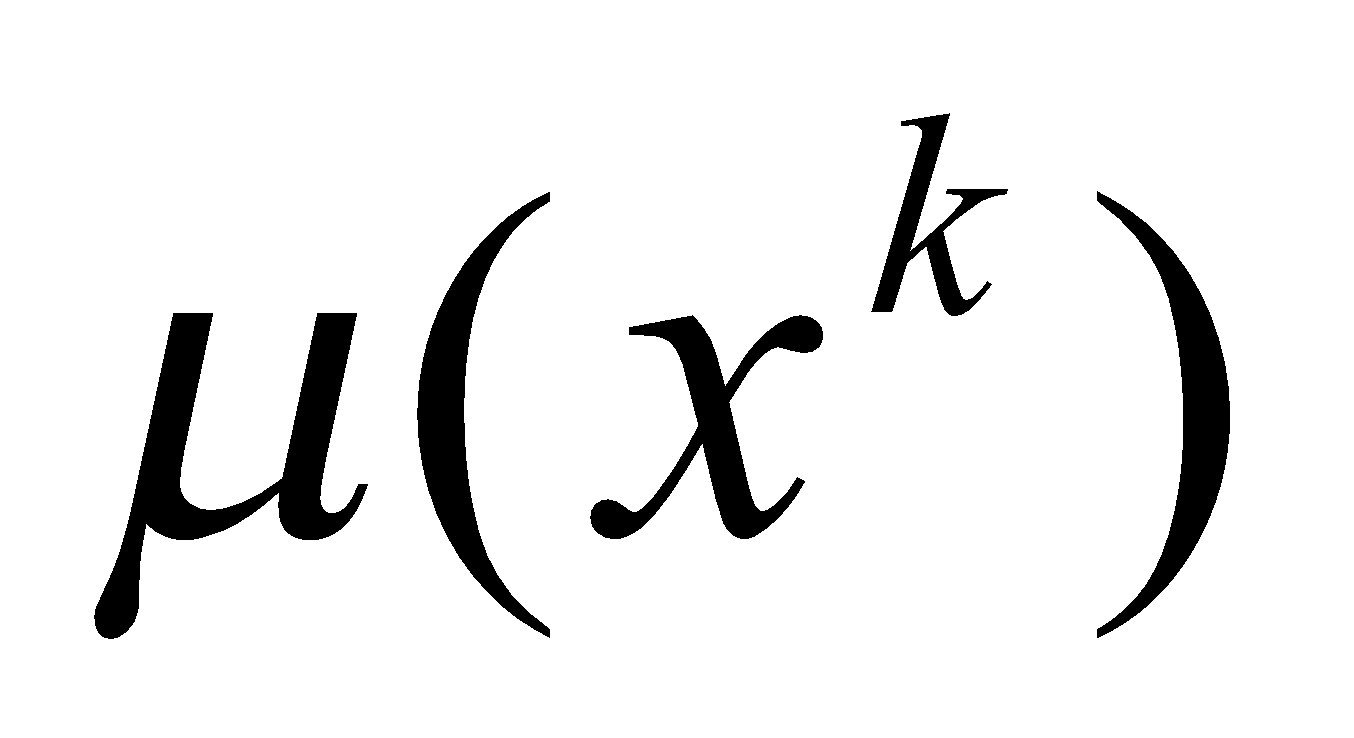
**Ш.1. Формирование исходной популяции.**

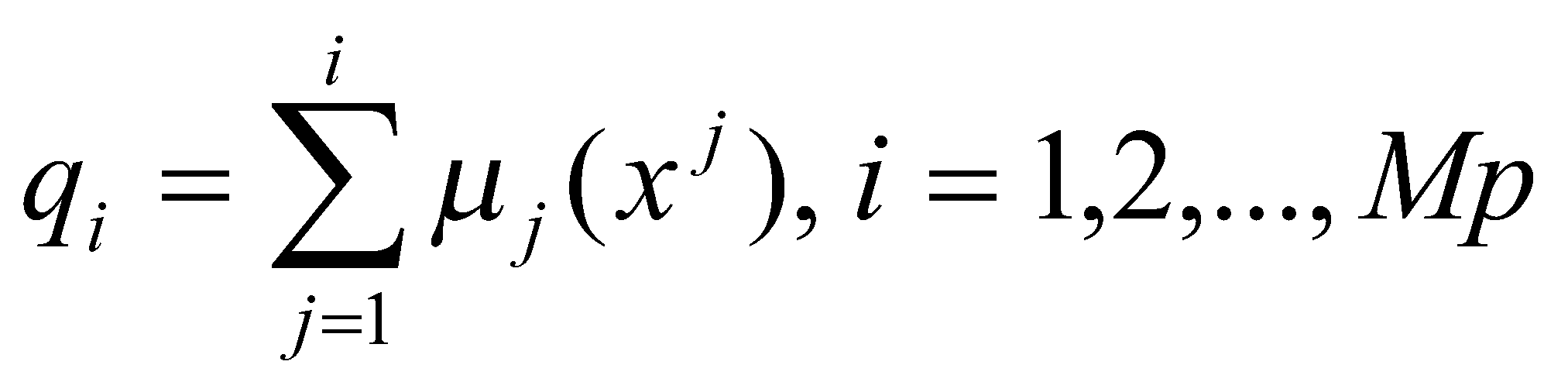
Ш.1.1.Задается номер популяции , максимальное количество популяций , номер итерации цикла , размер популяции .

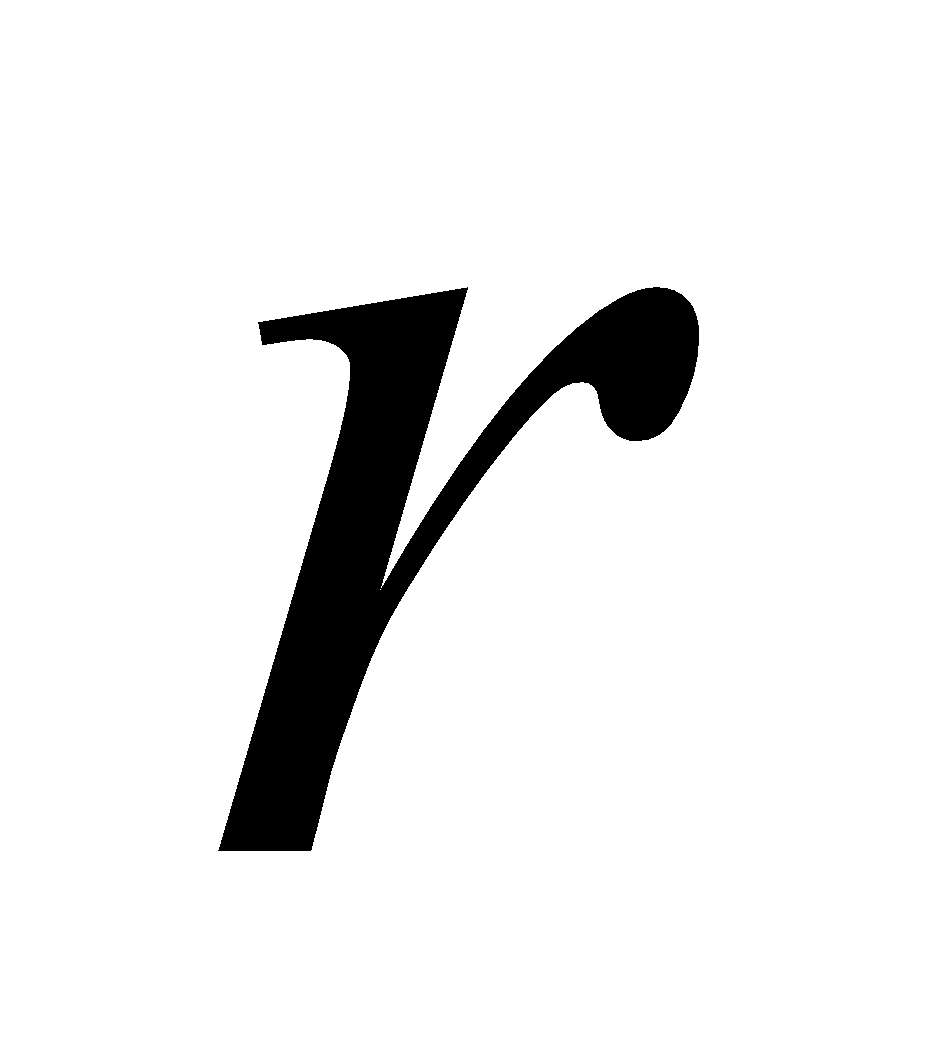
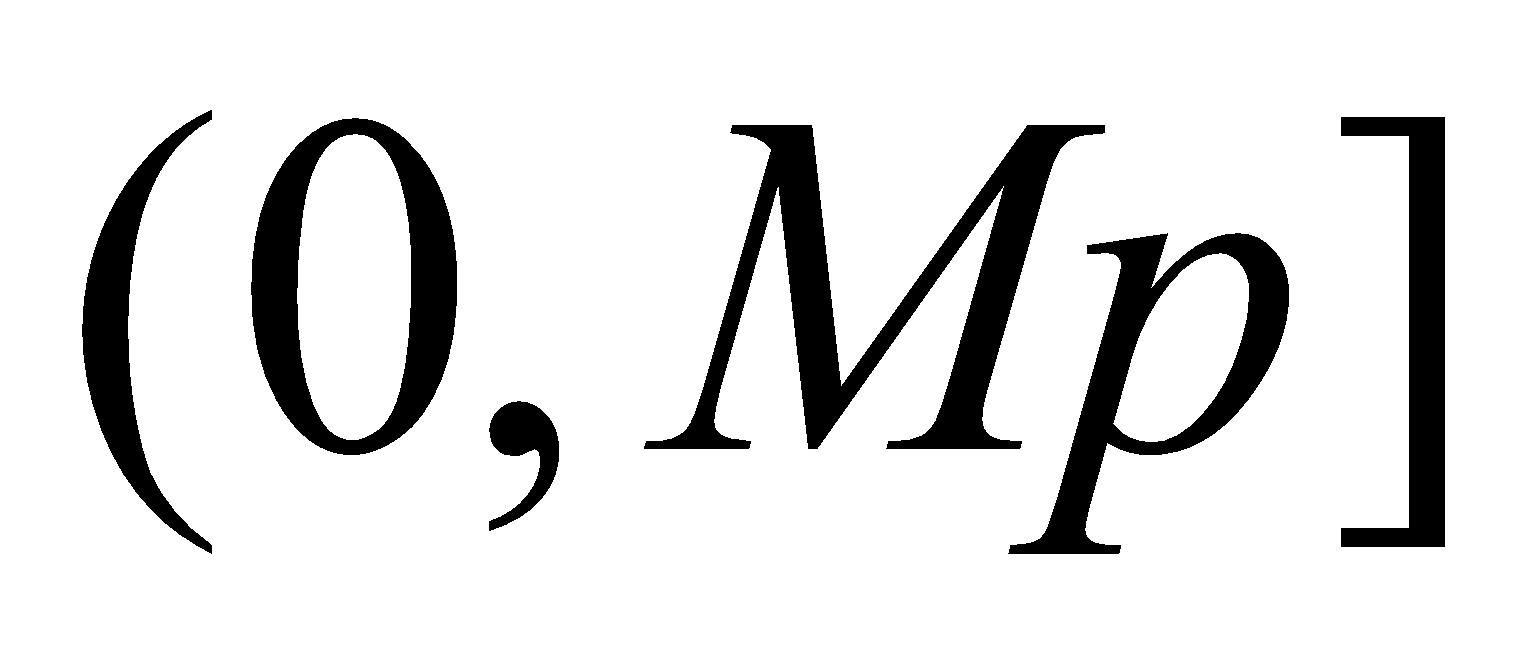
Ш.1.2. Случайным образом выбирается начальная точка  - исходная хромосома. Она может быть выбрана как внутренняя точка гиперкуба области допустимых значений . Из этой точки формируется исходная популяция. Для этого с помощью равномерного распределения на единичном отрезке   раз генерируется последовательность из  случайных точек . Используя линейное преобразование, каждая точка отображается на соответствующий ей промежуток . Составляя векторы из точек последовательности  при фиксированных , получаем  начальных векторов , координаты которых  имеют равномерное распределение на отрезках . Таким образом может быть сформирована начальная популяция .

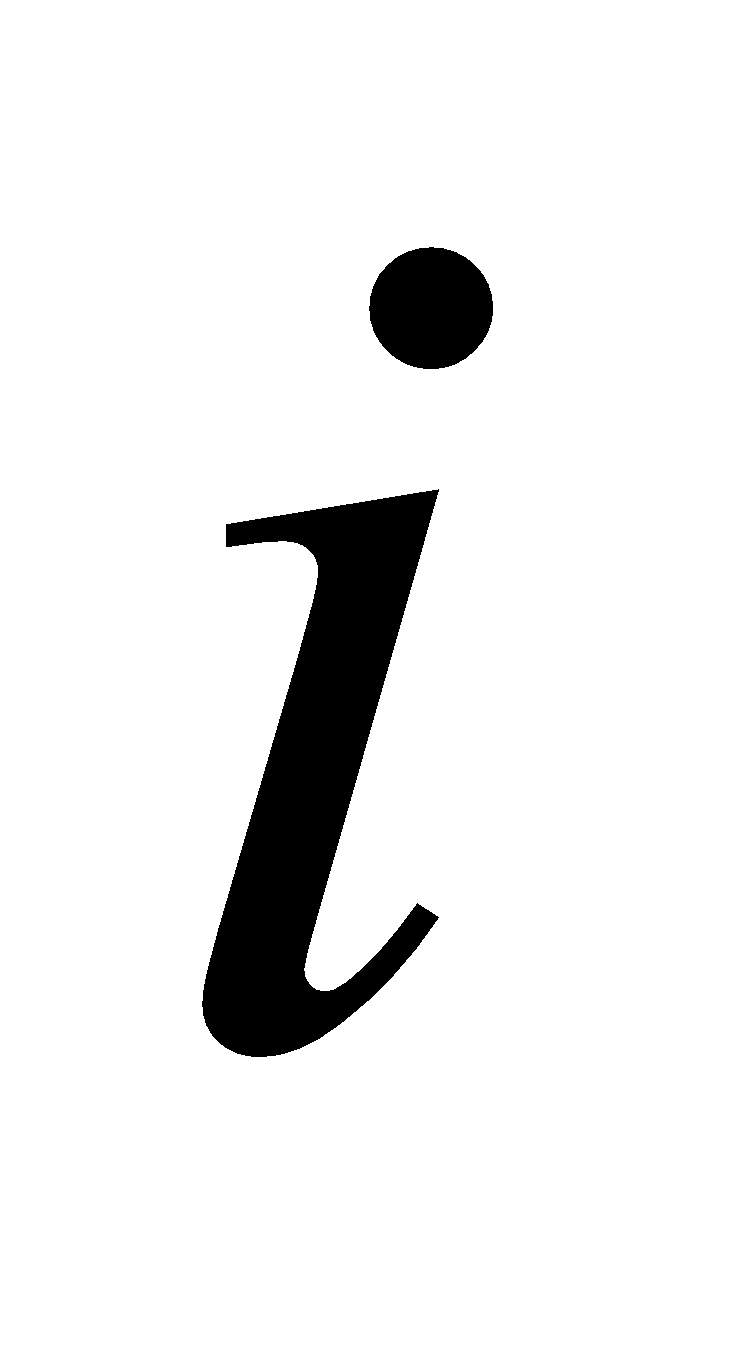
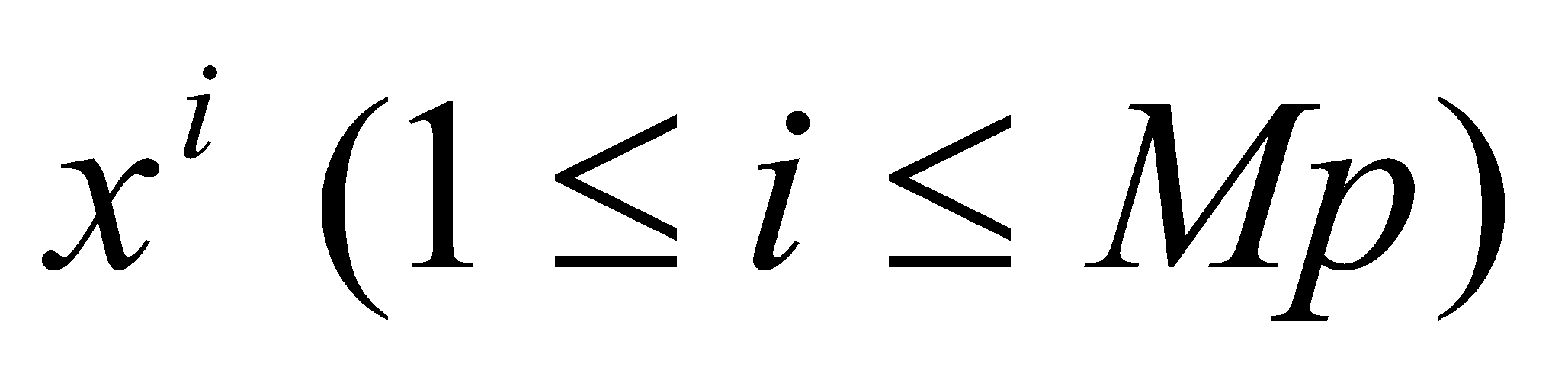
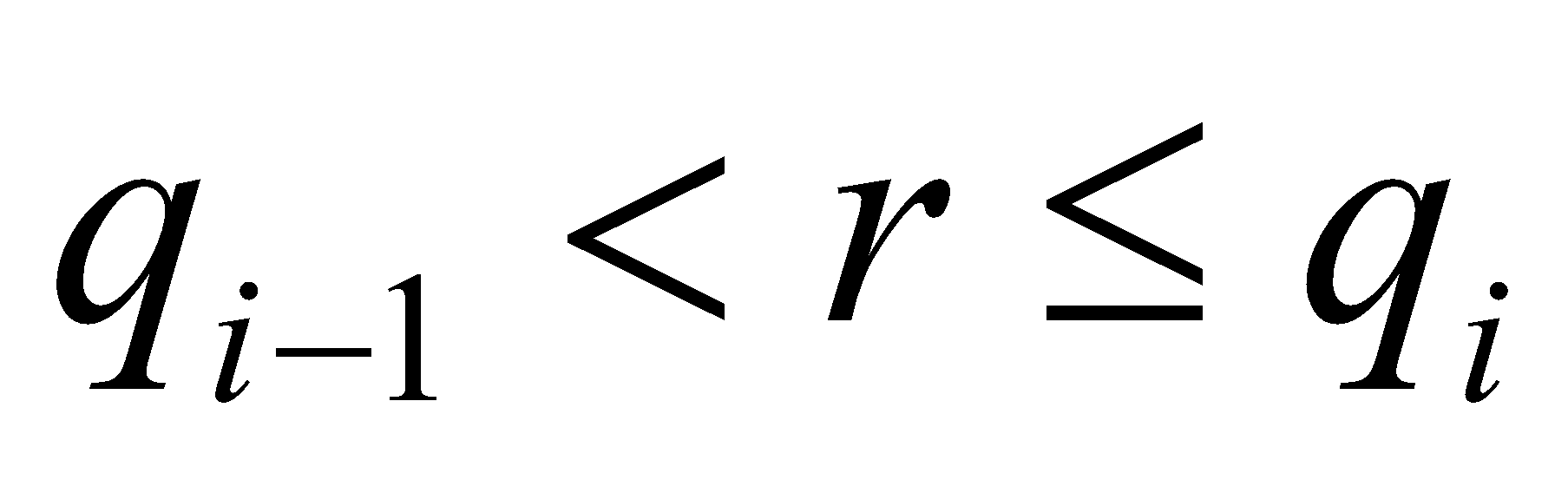
Ш.1.3. Вычисляется значение функции фитнеса для каждой особи  и популяции  в целом .

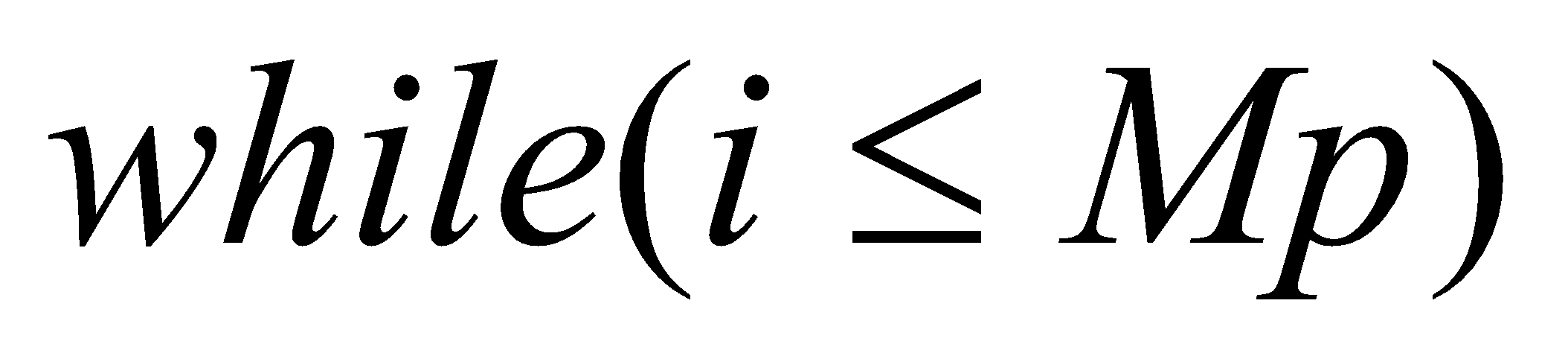
**Ш.2. Отбор (селекция).**

*Селекция* – это операция, которая осуществляет отбор особь (хромосом)  в соответствии со значениями функции фитнеса  для последующего их скрещивания.

Ш.2.1. Вычислить кумулятивную вероятность .

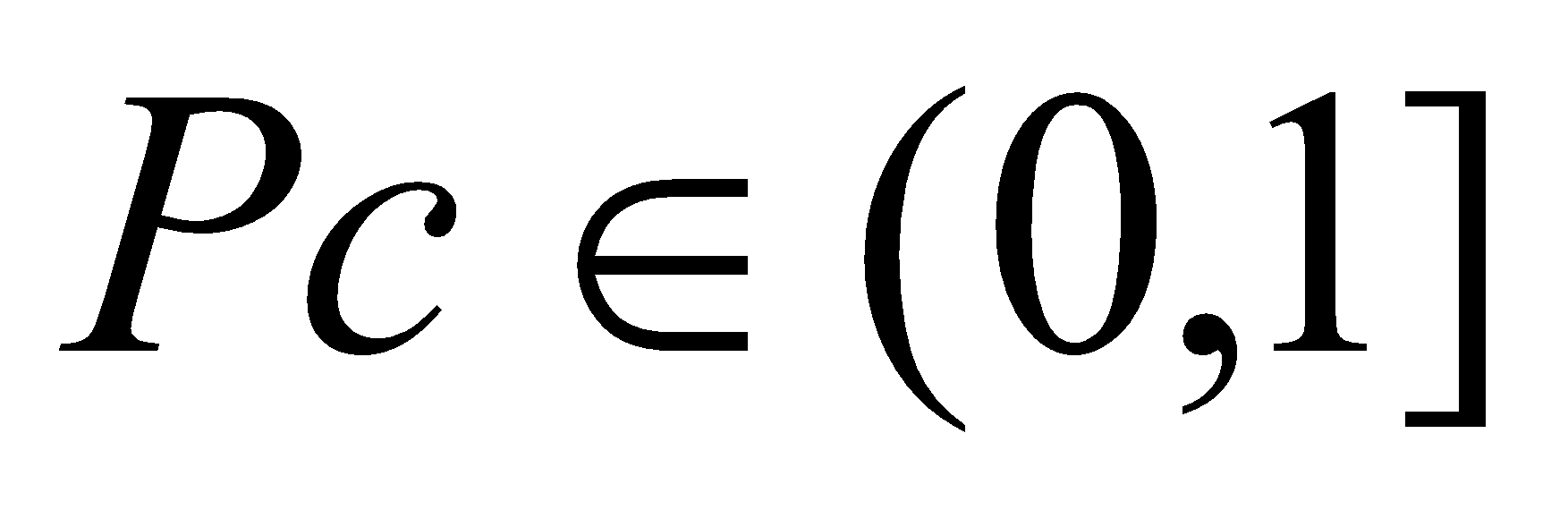
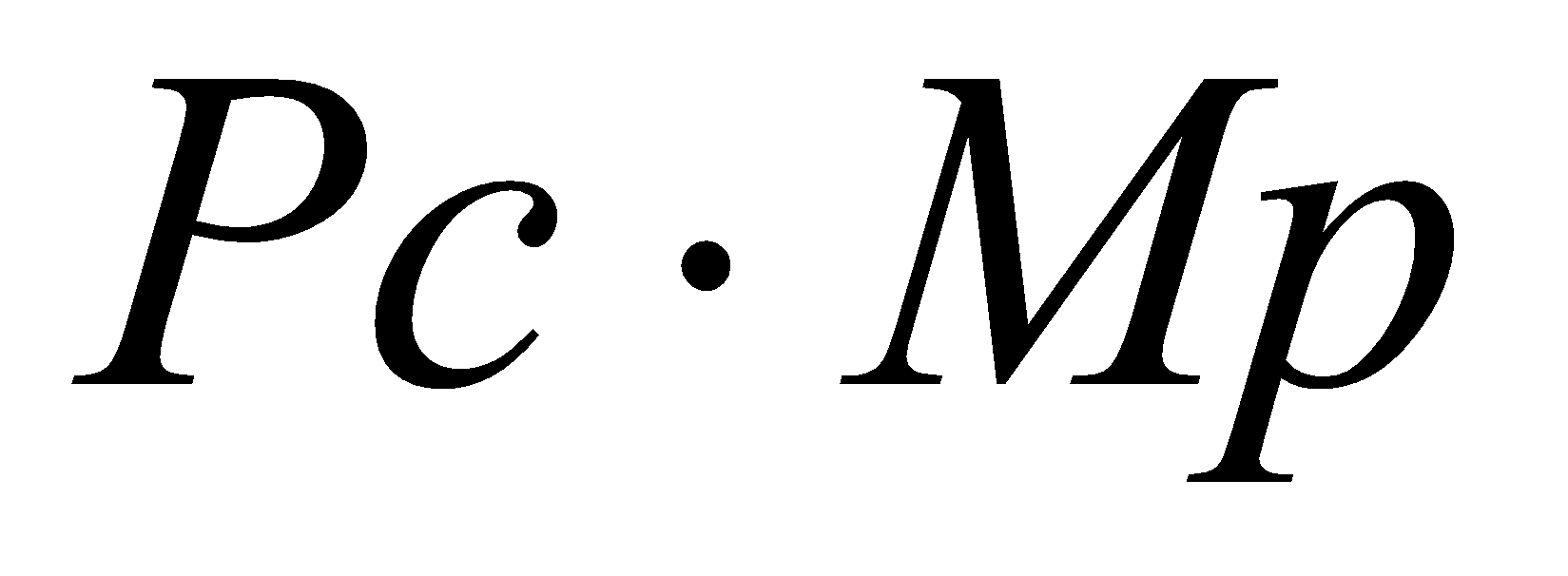
Ш.2.2. Сформировать случайное действительное число  в интервале .

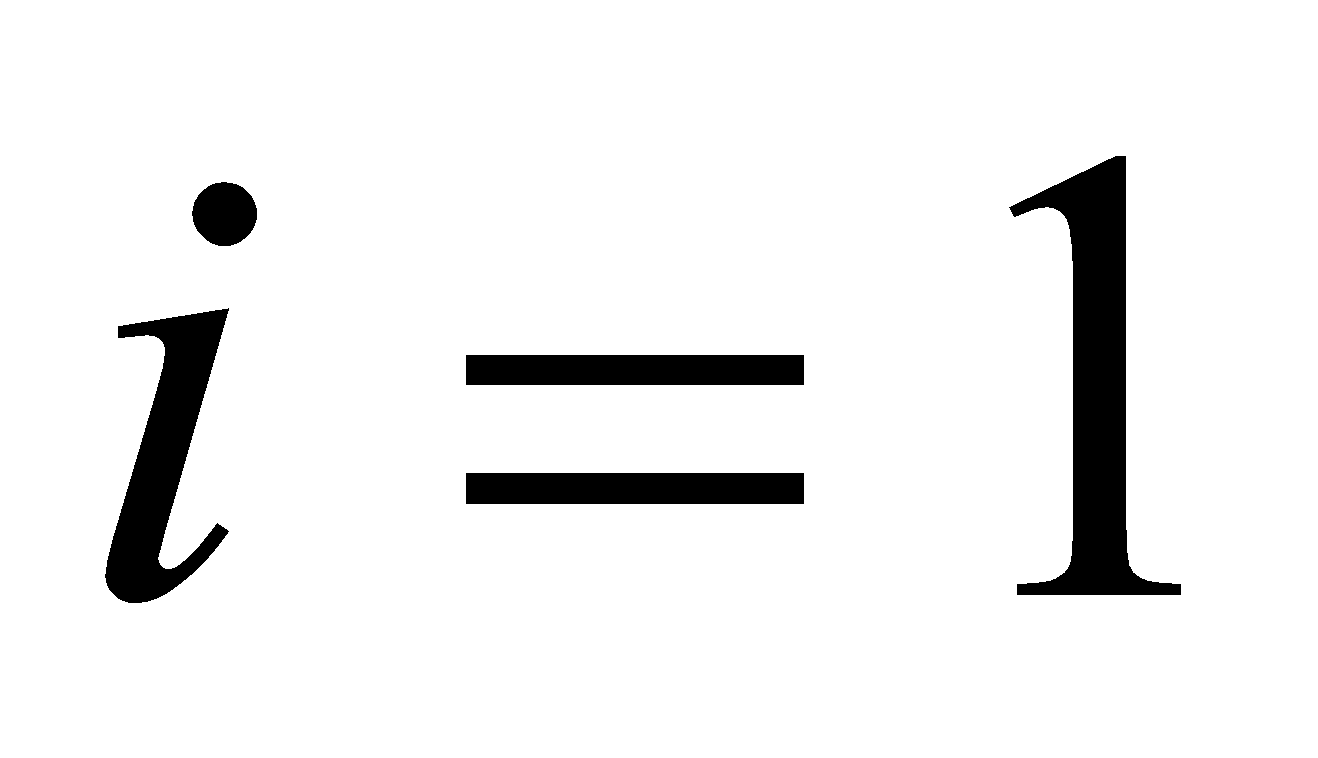
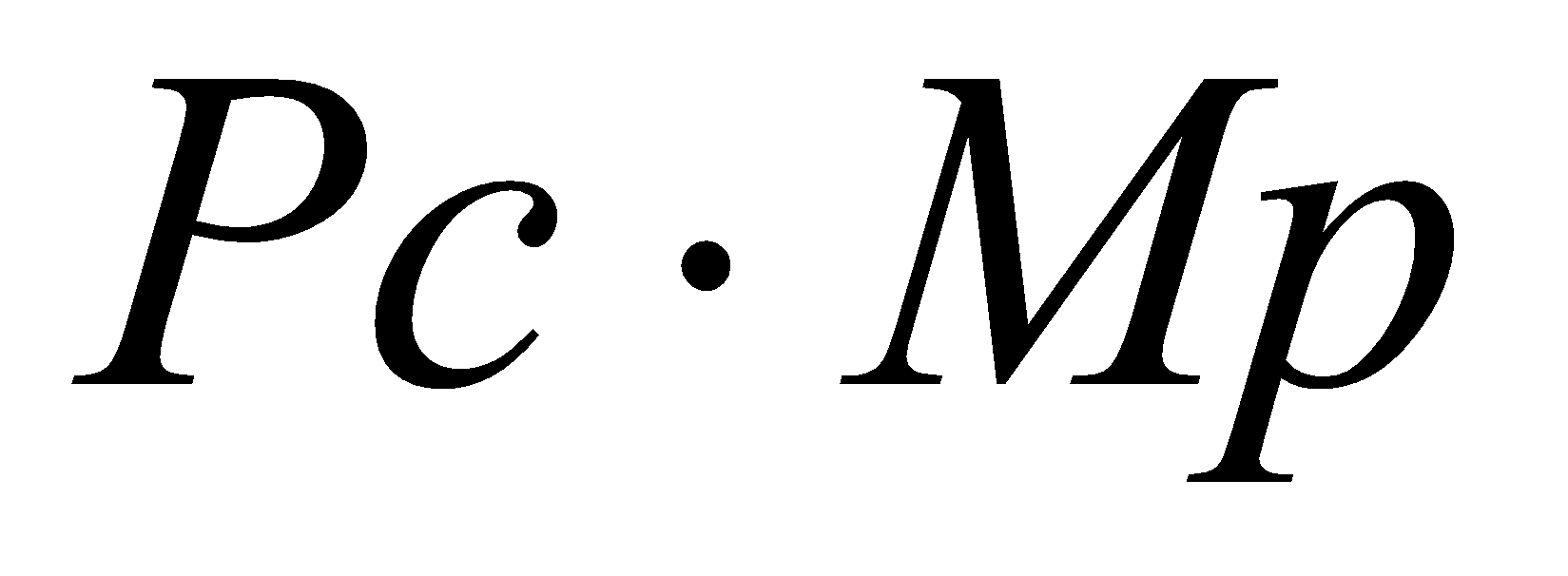
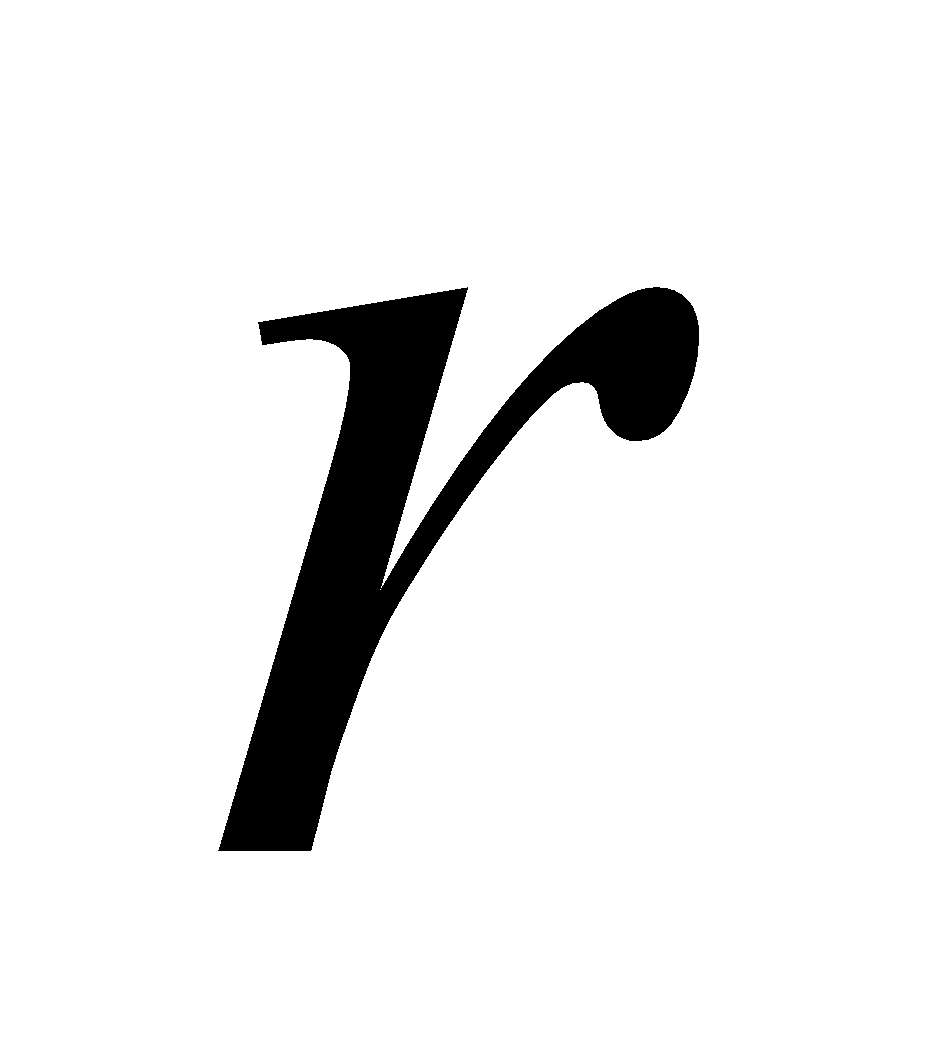
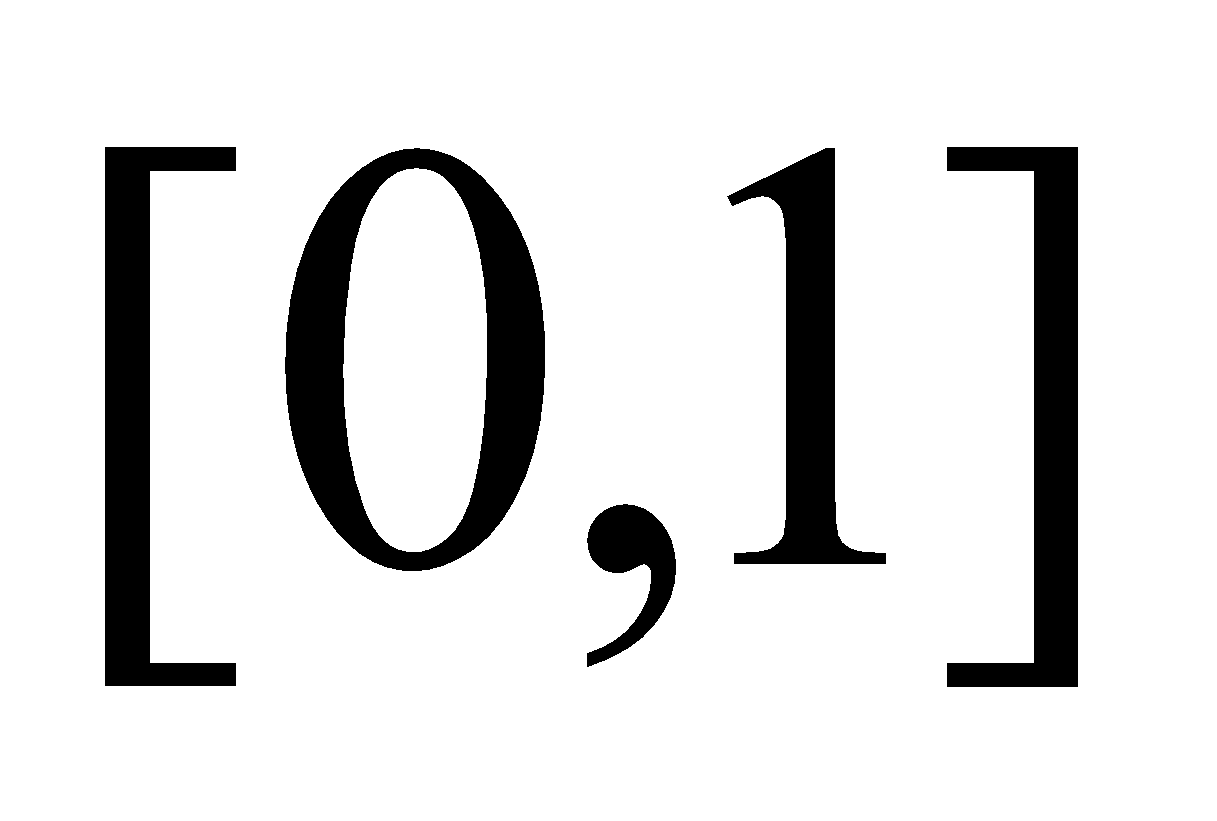
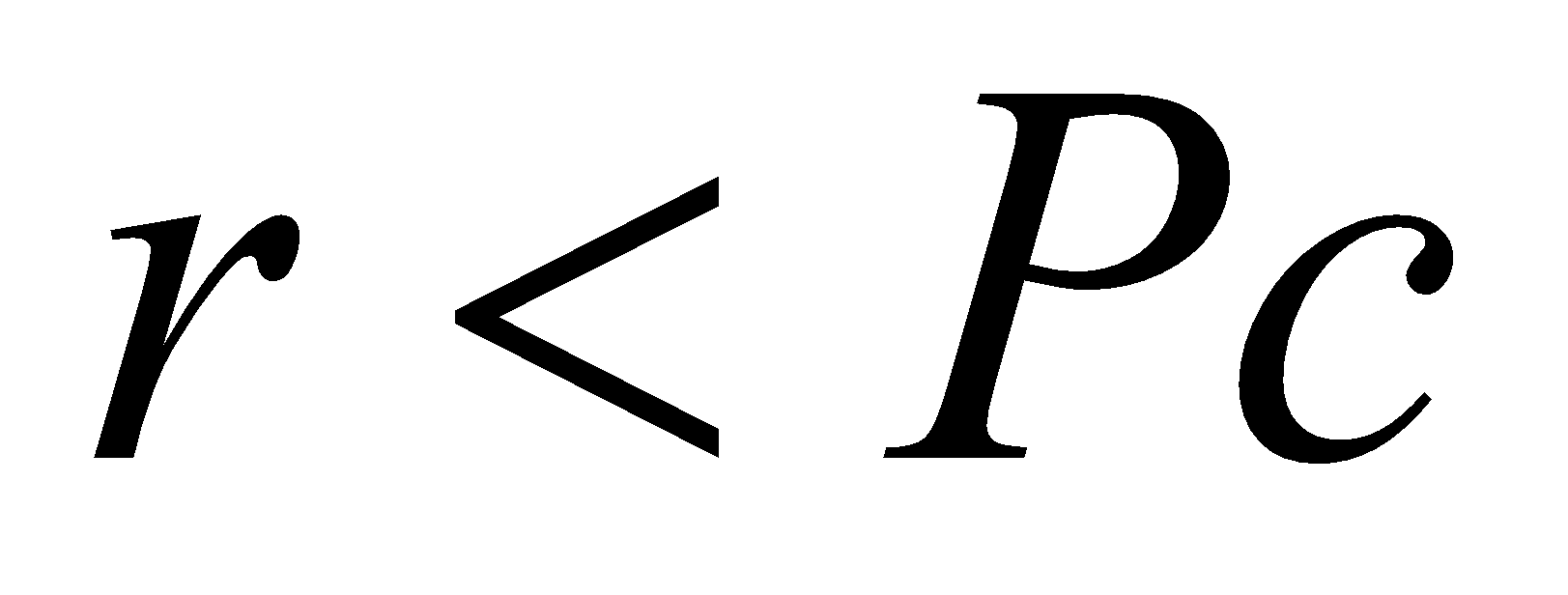
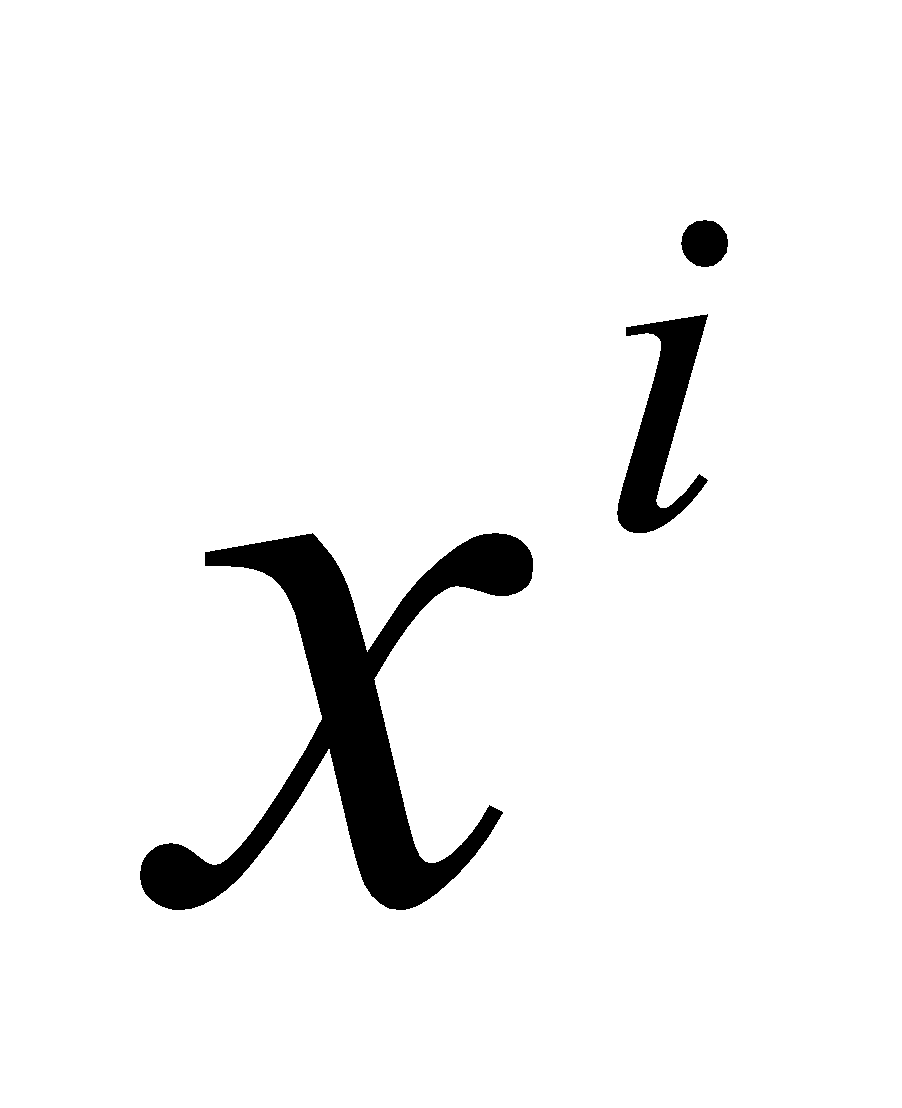
Ш.2.3. Выбрать  - ю хромосому  так, чтобы .

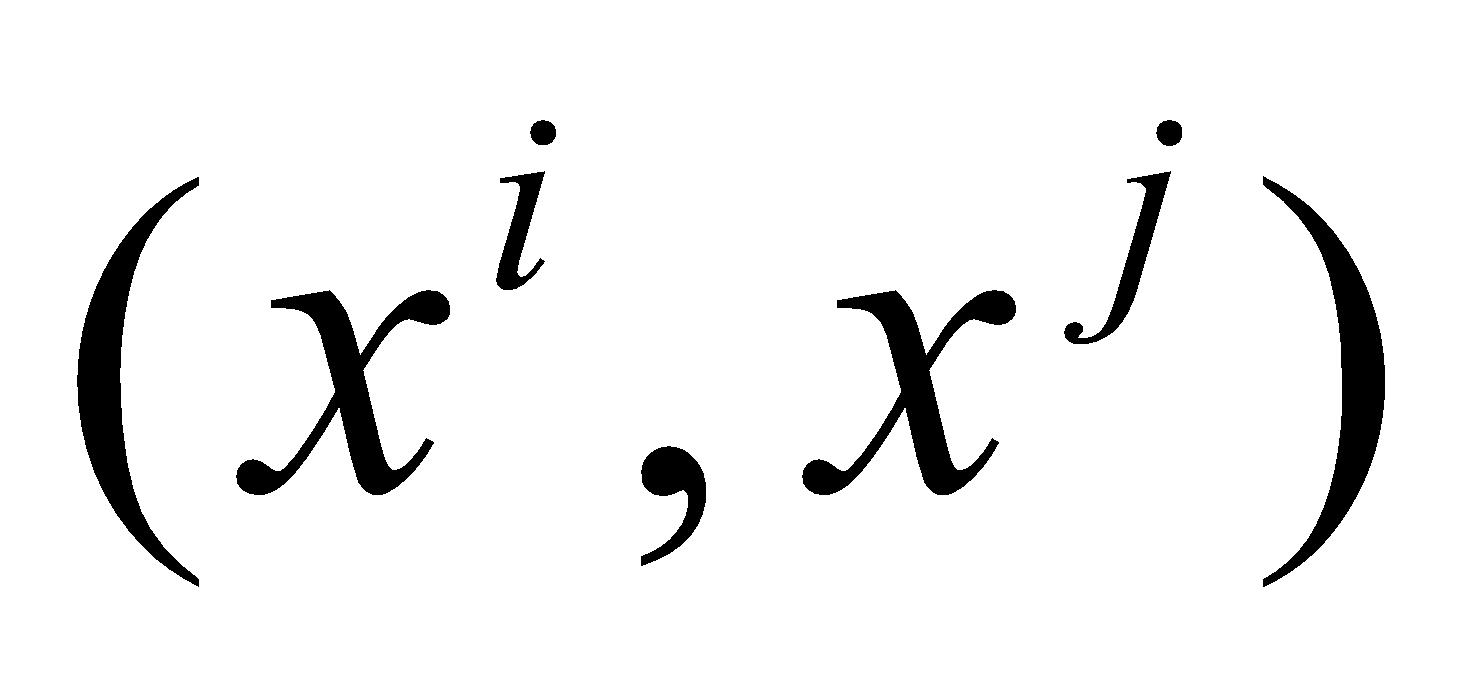
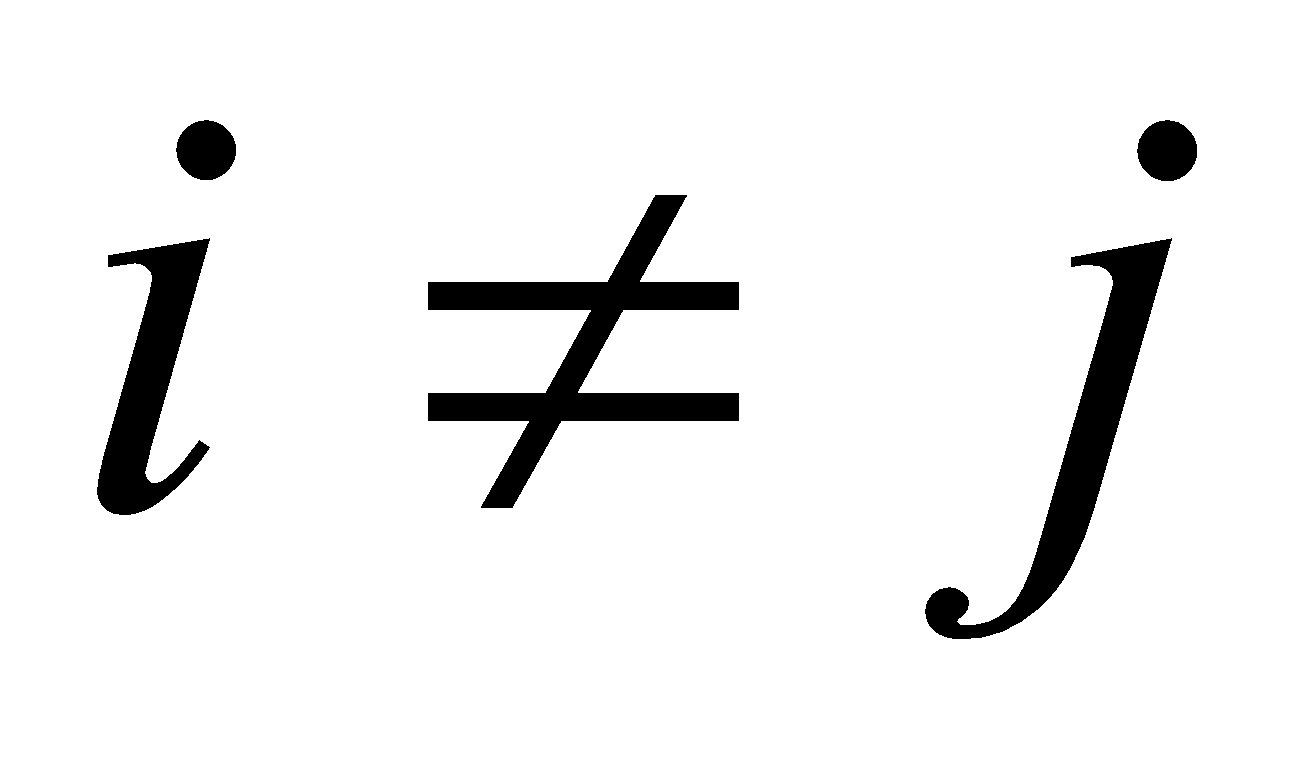
Ш.2.4. Перейти на Ш.2.2. до тех пор, пока не будет сформирована новая популяция ().

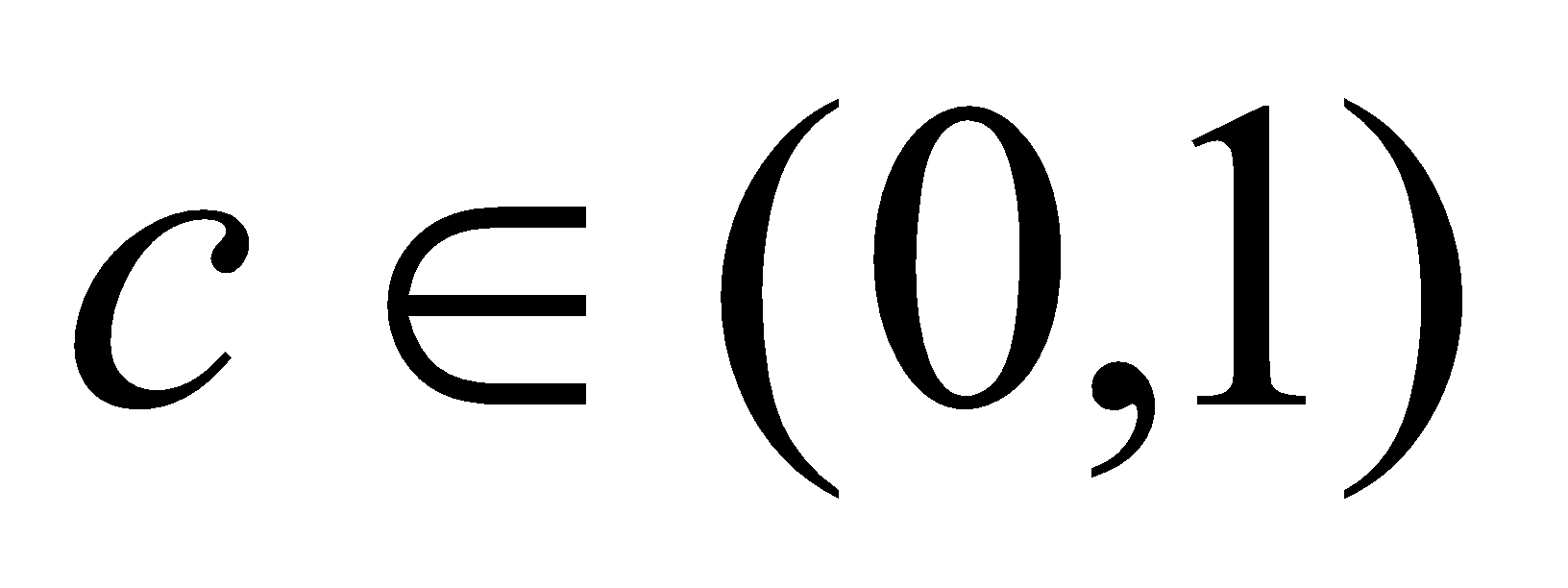
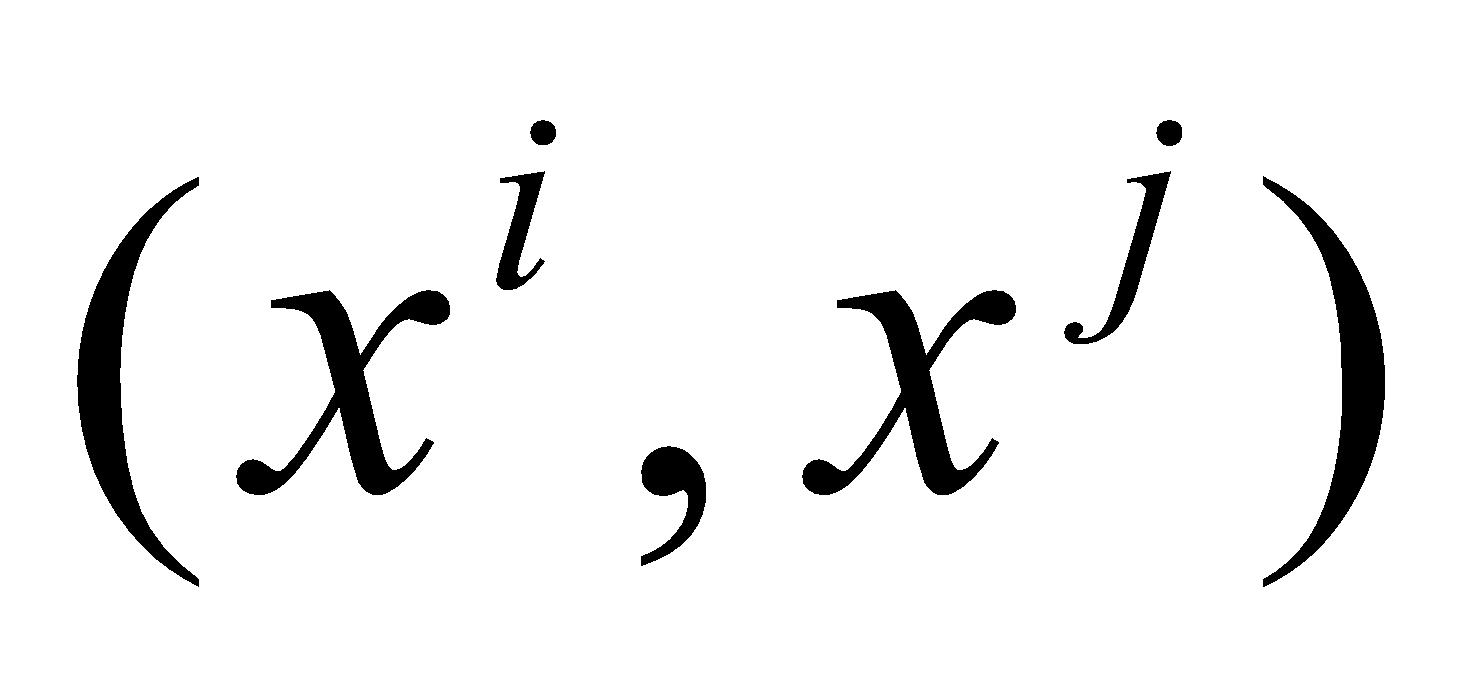
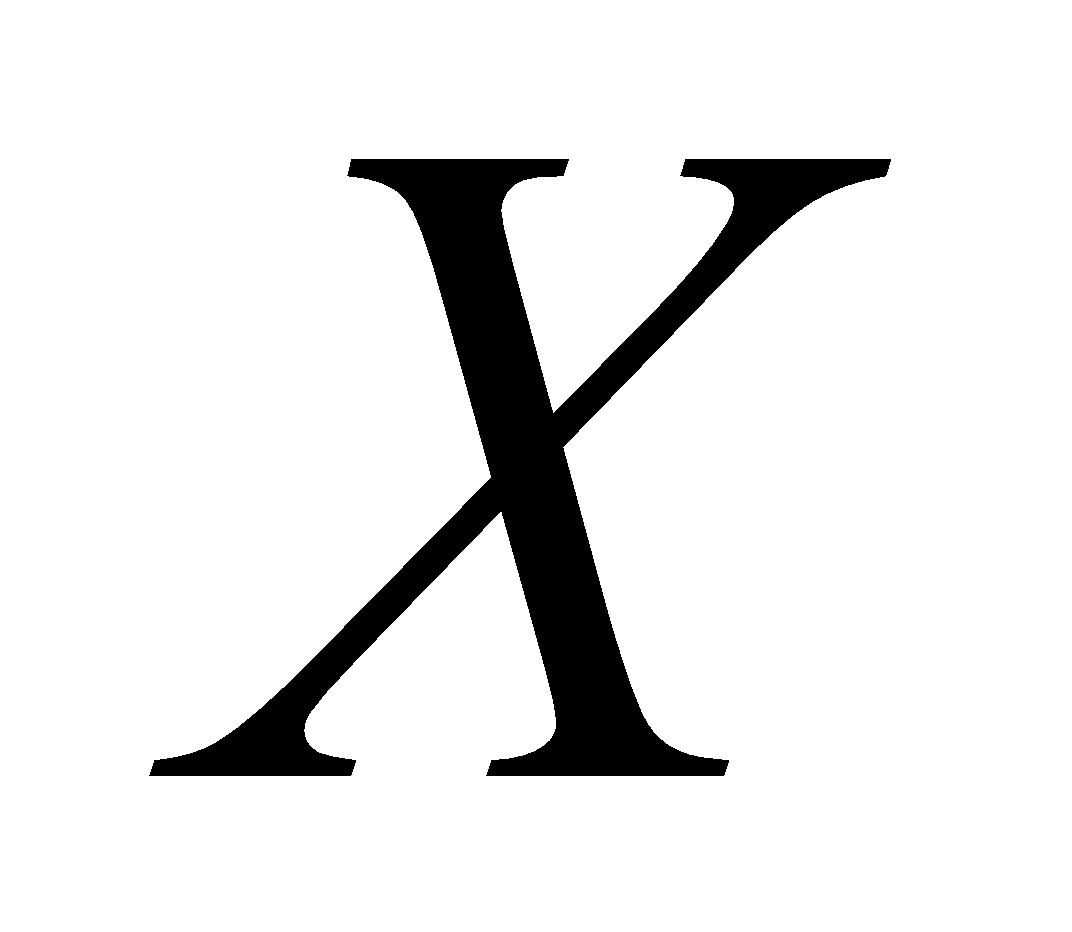
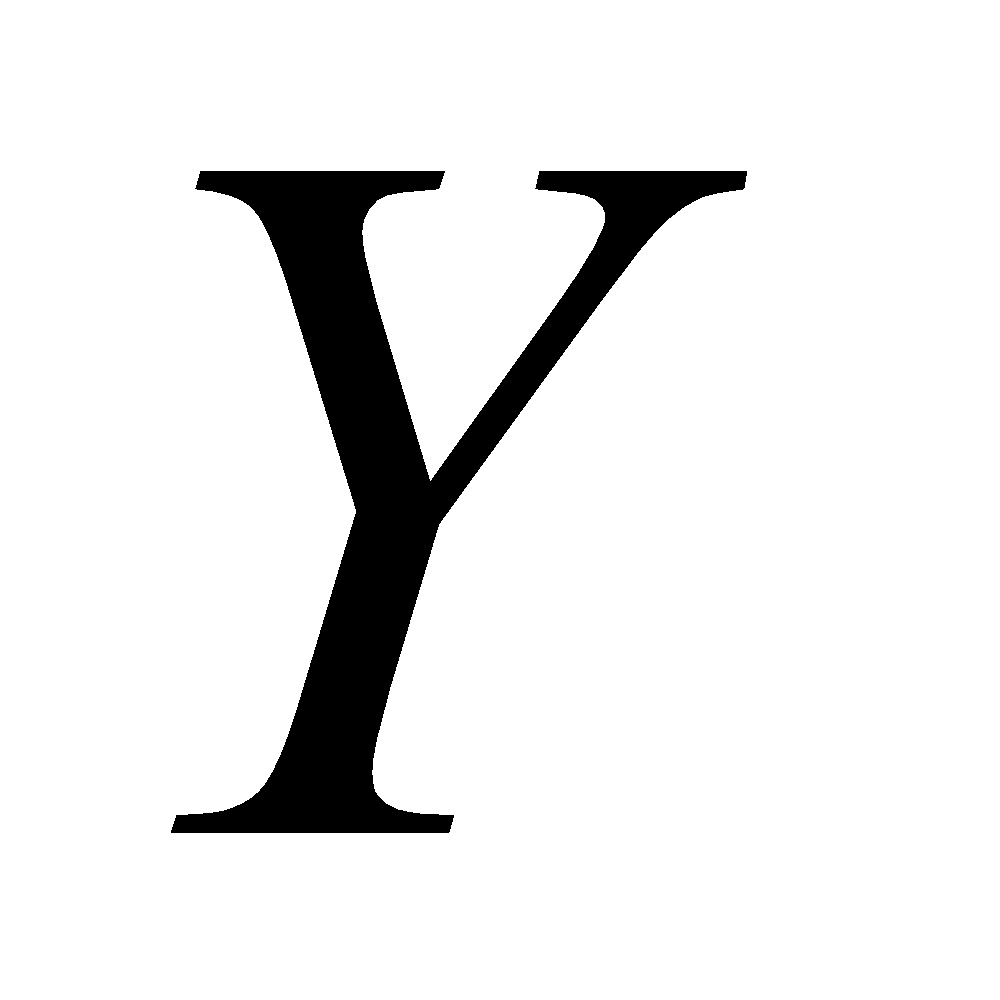
**Ш.3. Кроссинговер (скрещивание).**

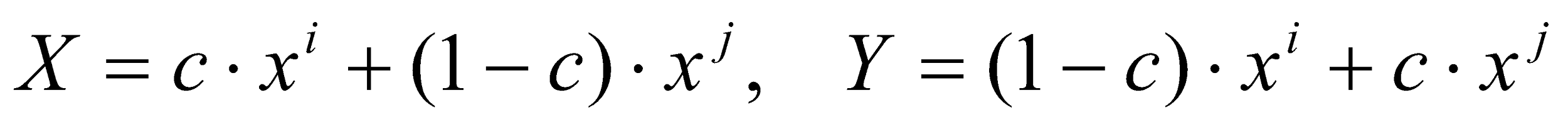
Скрещивание – это операция, при которой из нескольких, обычно двух хромосом (особей), называемых родителями, порождается одна или несколько новых, называемых потомками.

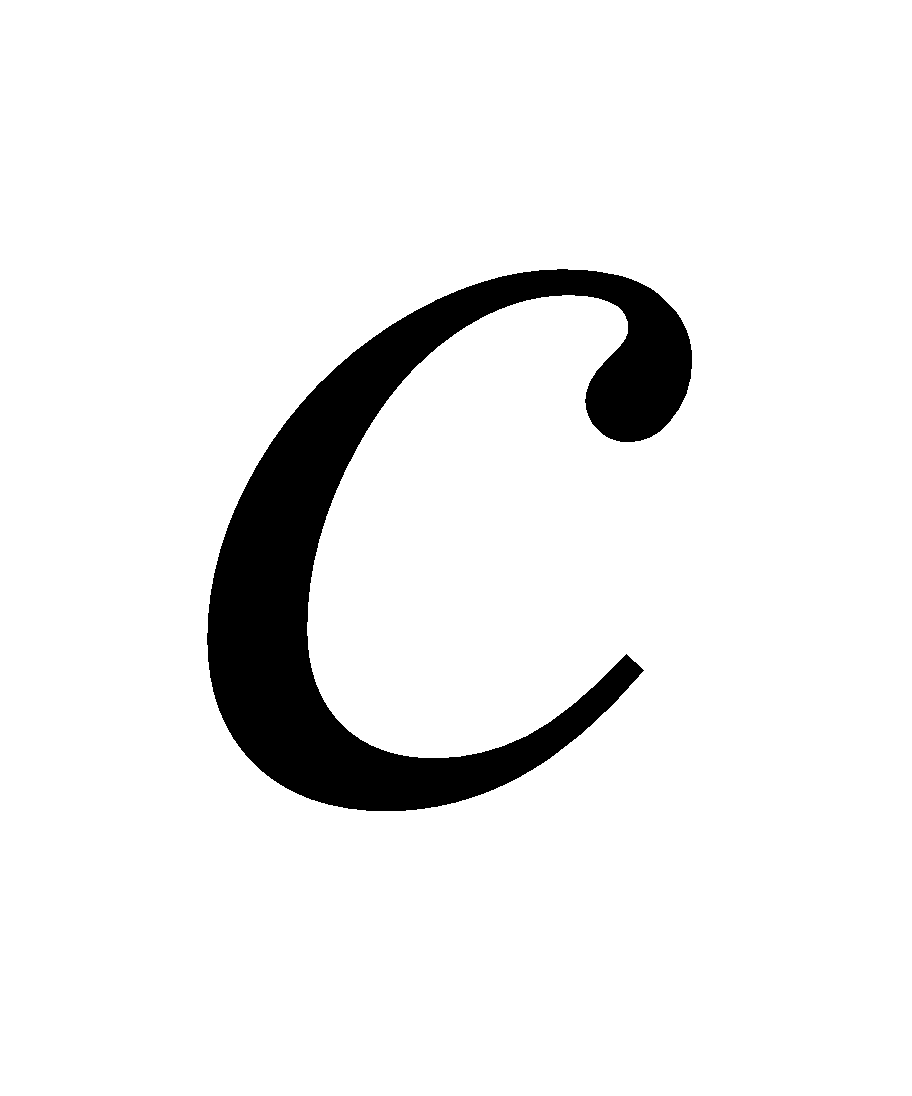
Ш.3.1. Определяется параметр  как вероятность кроссинговера. Эта вероятность дает ожидаемое число  хромосом, подвергаемых операции кроссинговера.

Ш.3.2. Для операции кроссинговера выполняется процесс, повторяющийся от  до : формируется случайное действительное число из сегмента , при этом, если , то хромосома  выбирается как родительская.

Ш.3.3. Отбираются пары родительских хромосом , где . Действие оператора кроссинговера осуществляется следующим образом:

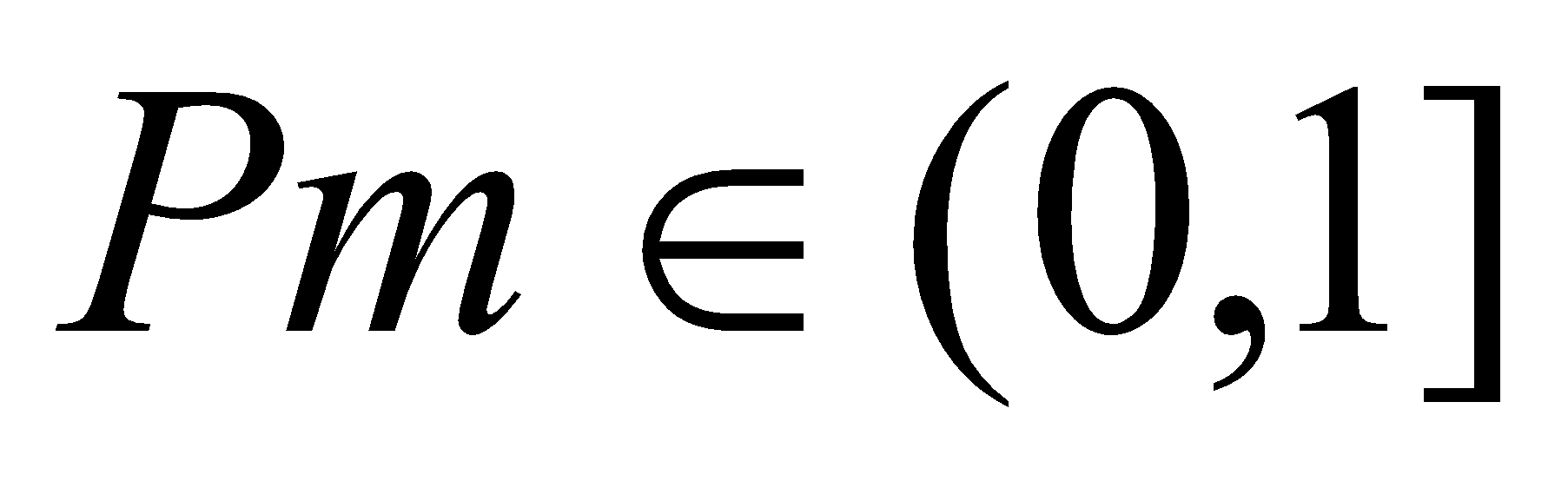
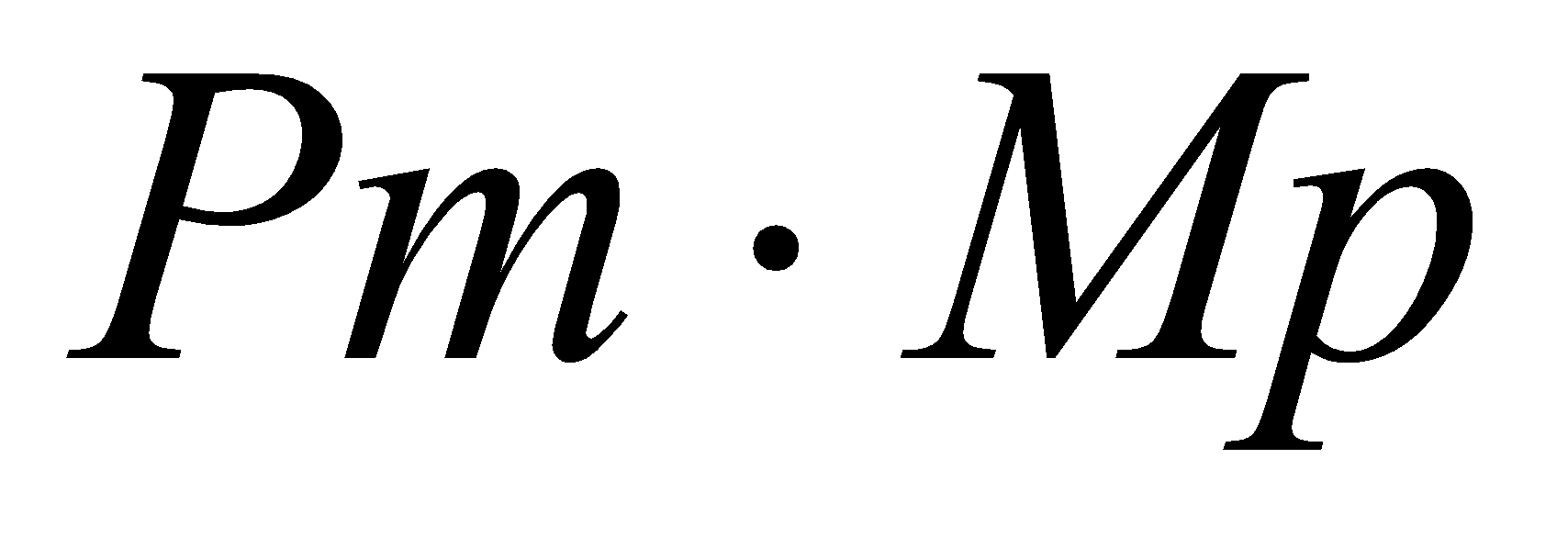
Ш.3.4. Формируется случайное число , затем оператор кроссинговера, действующий на исходные пары , производит две хромосомы потомки  и :

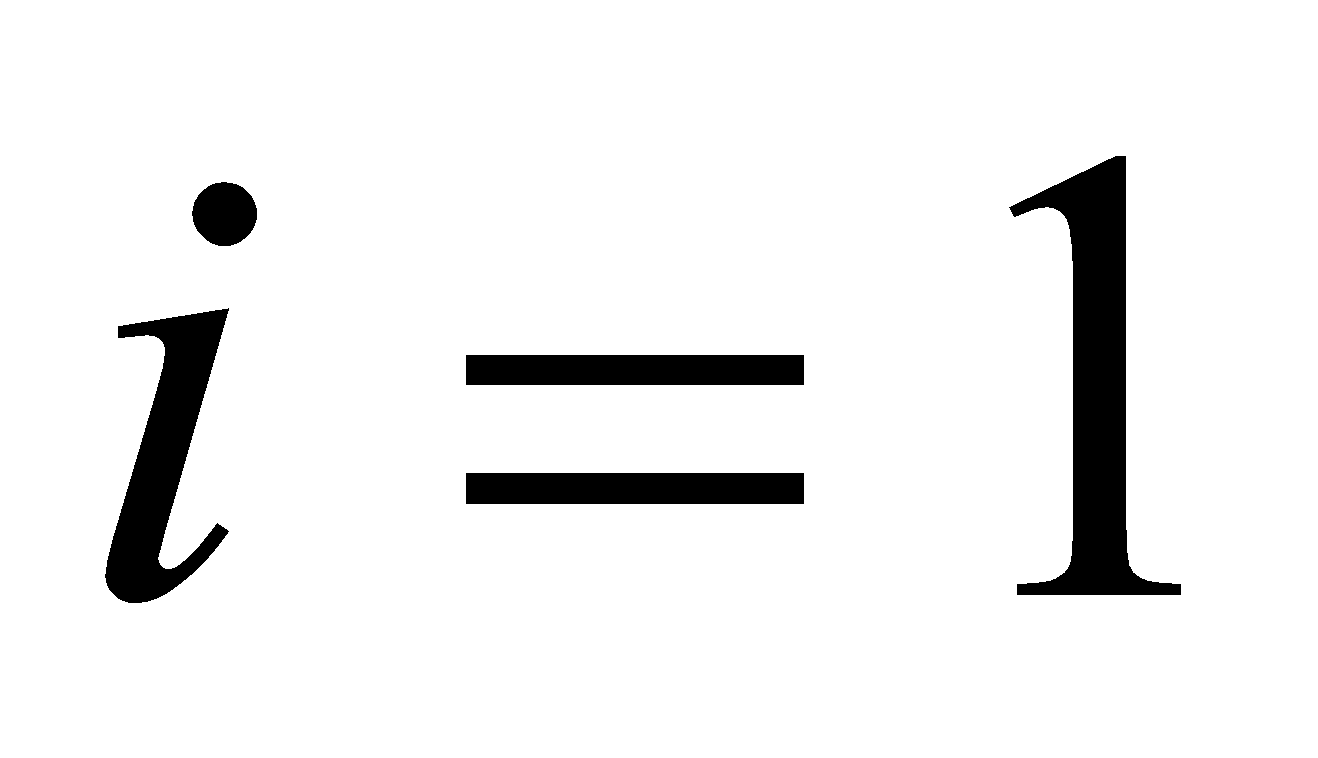
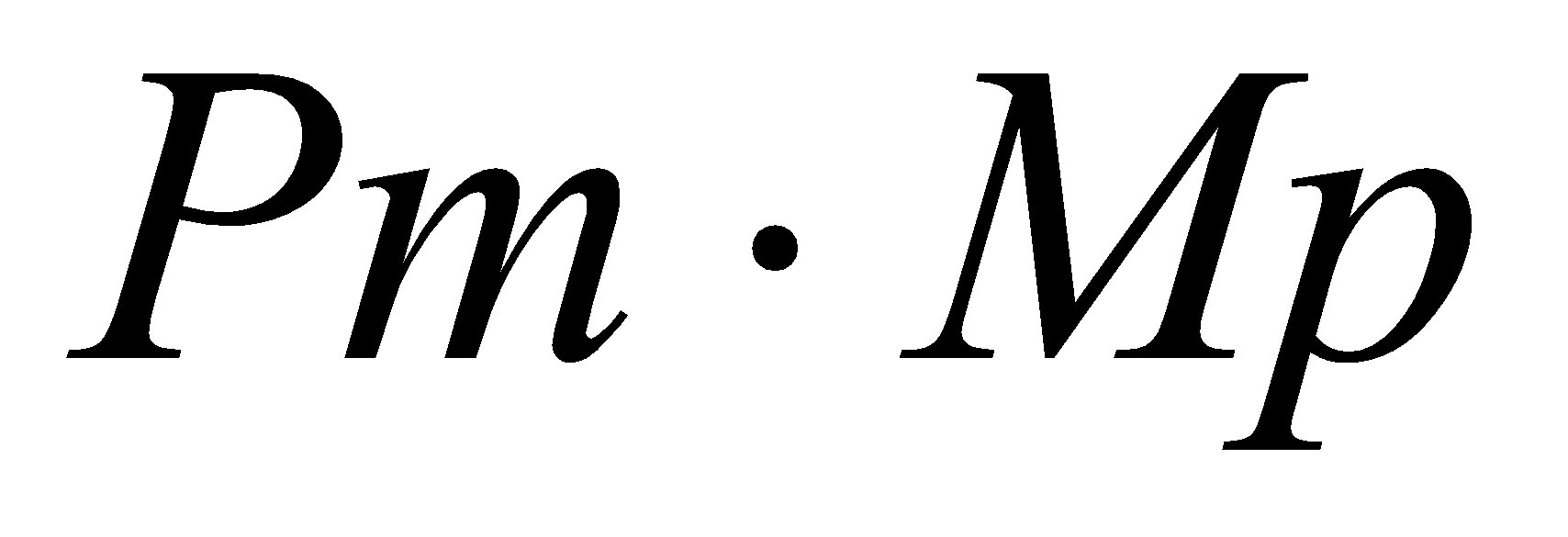
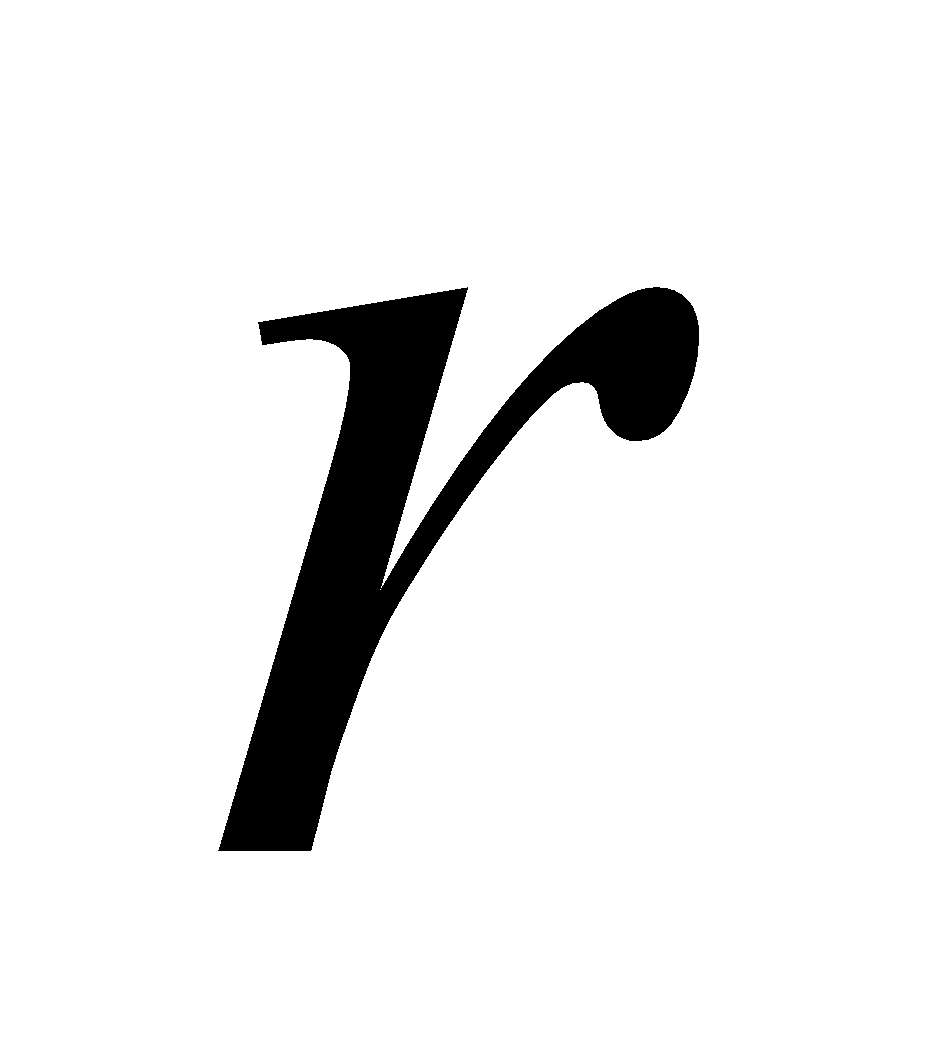
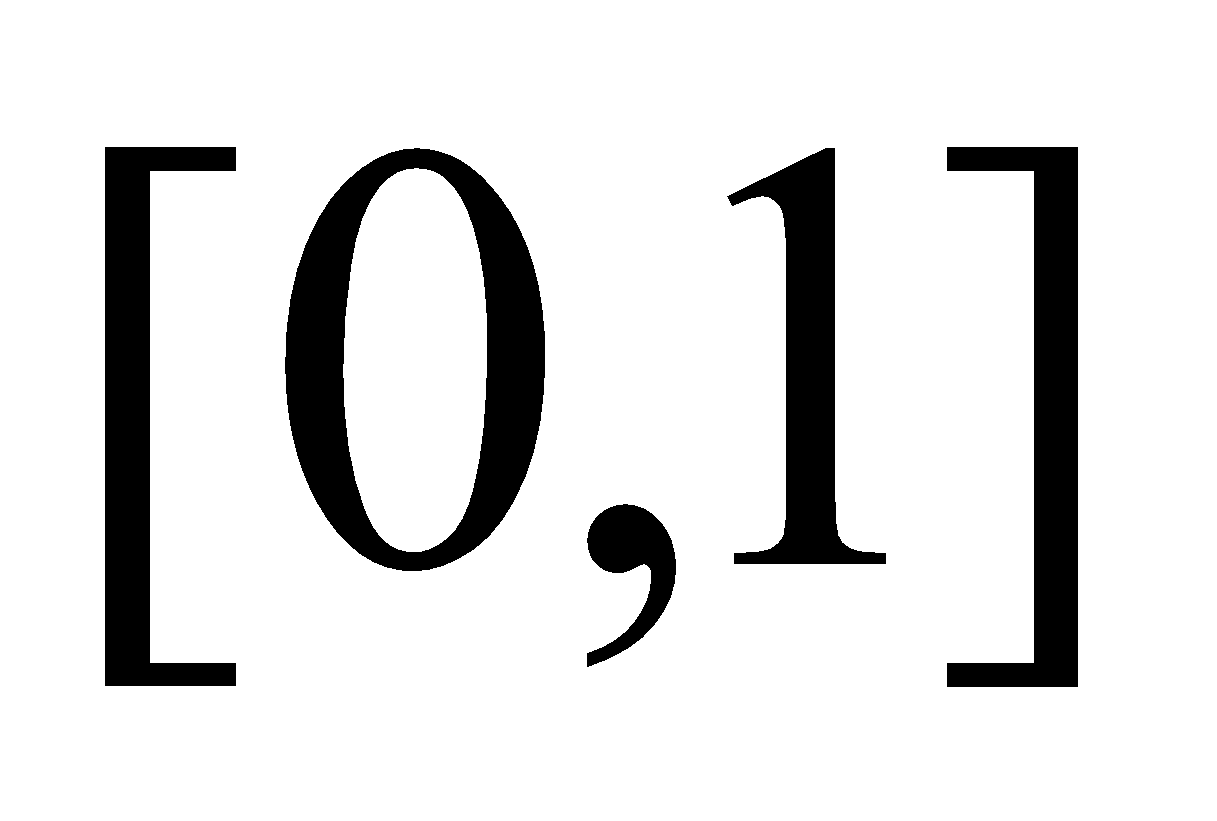
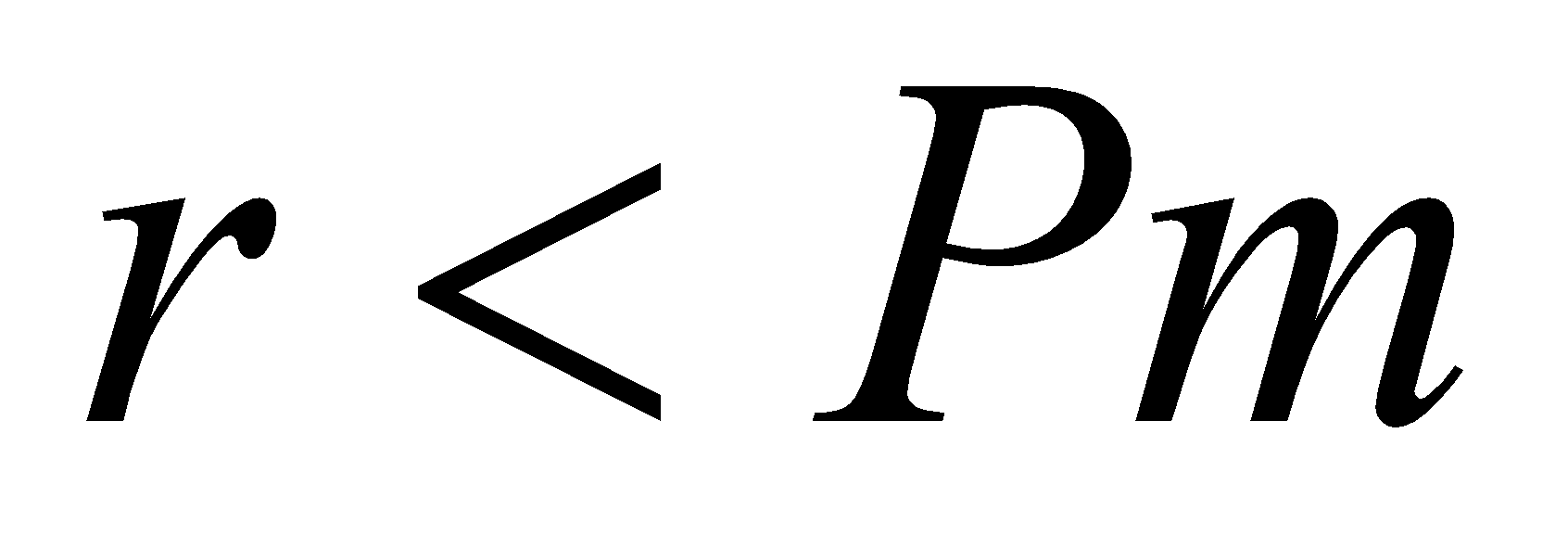
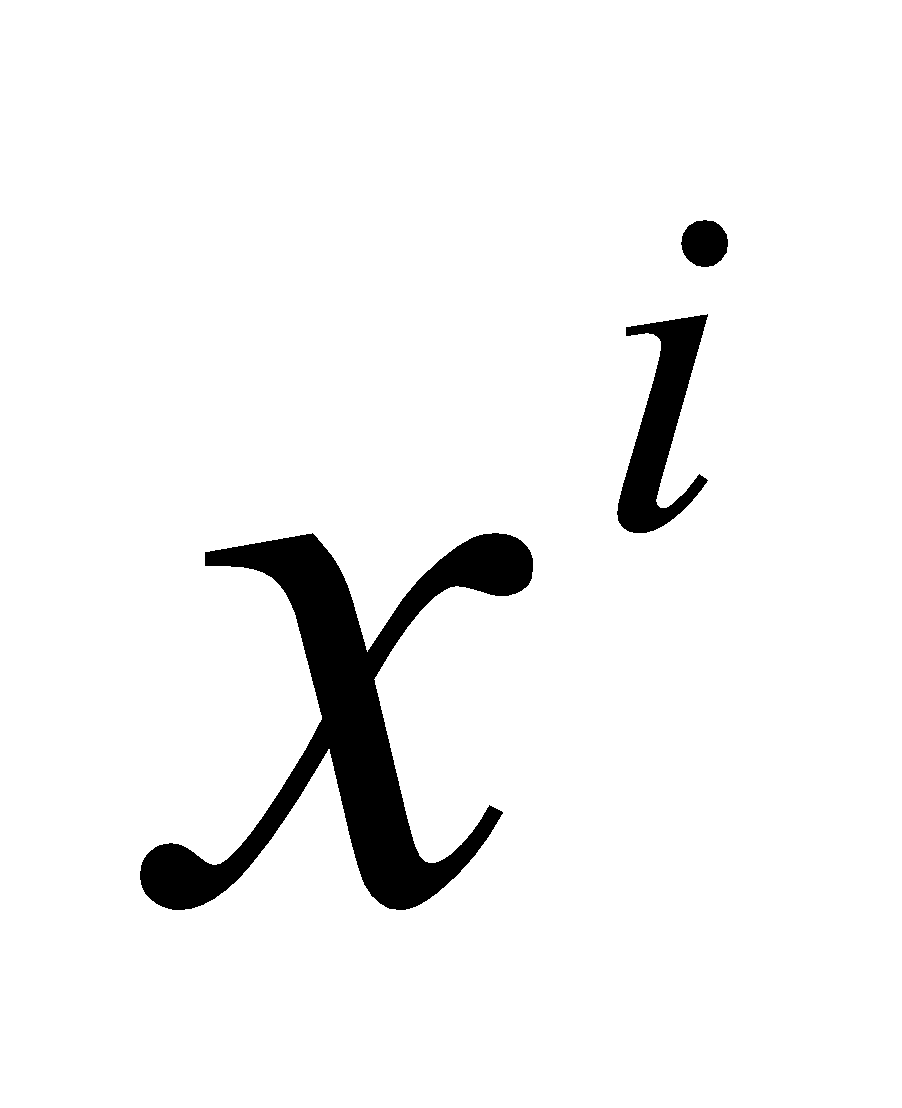
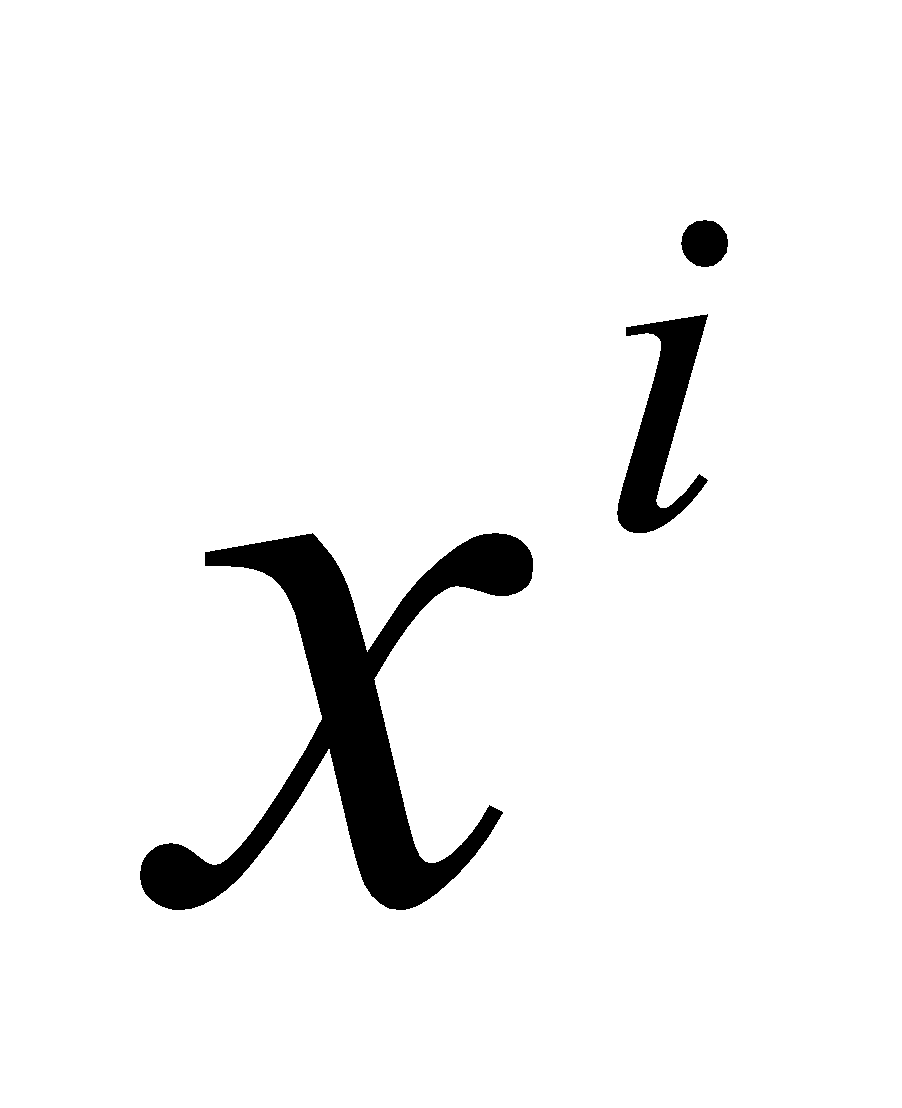
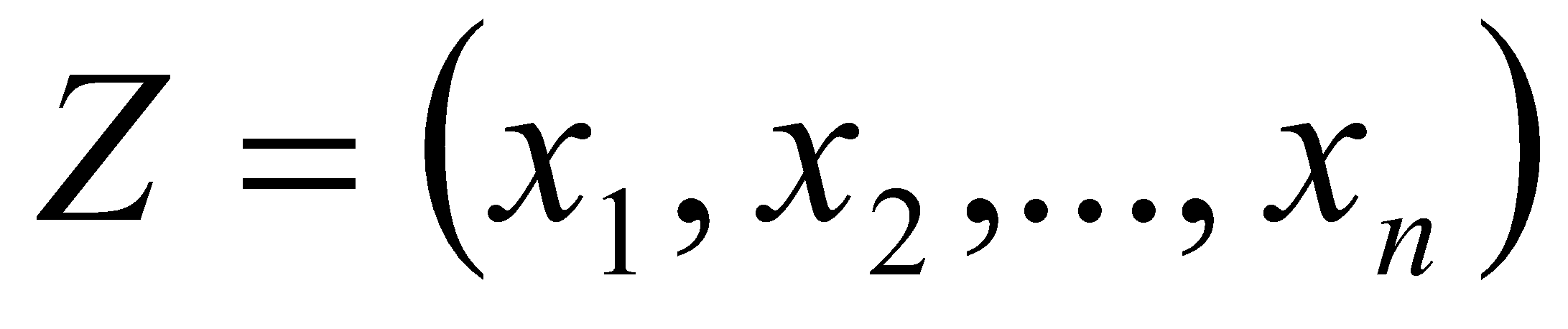
.

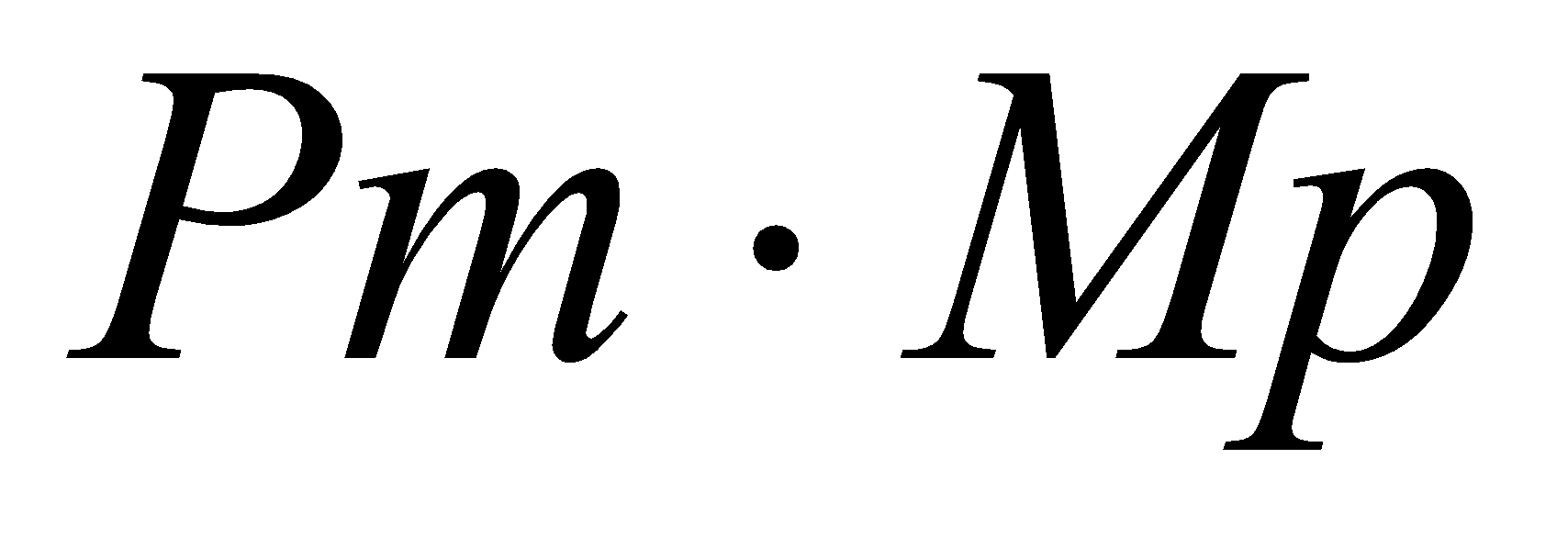
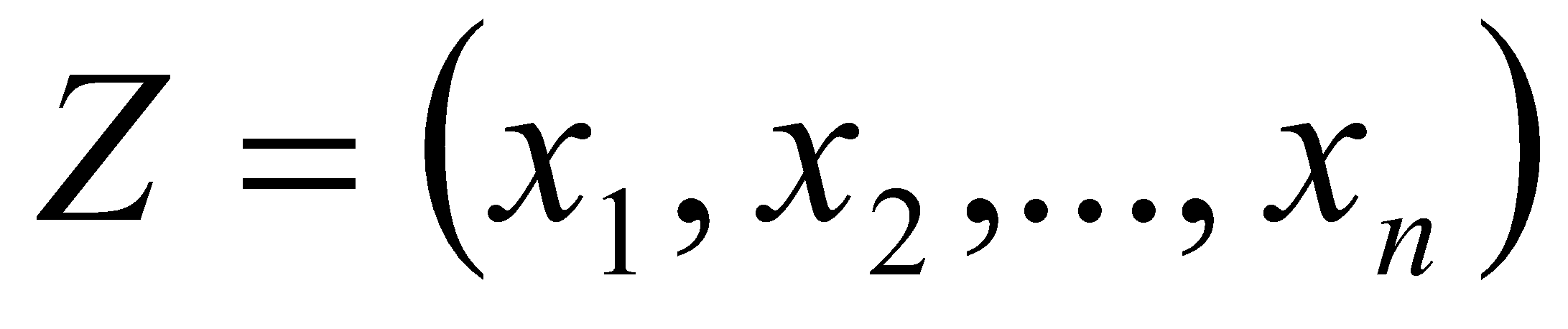
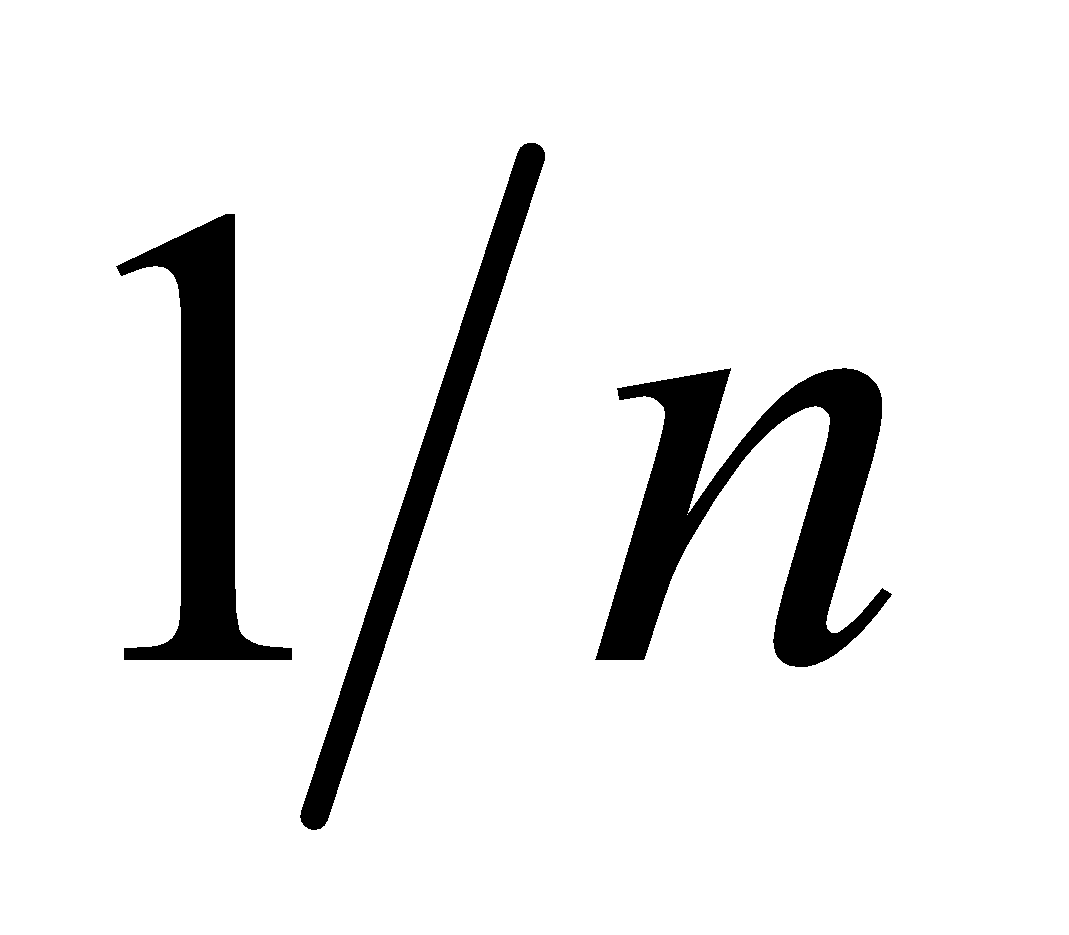
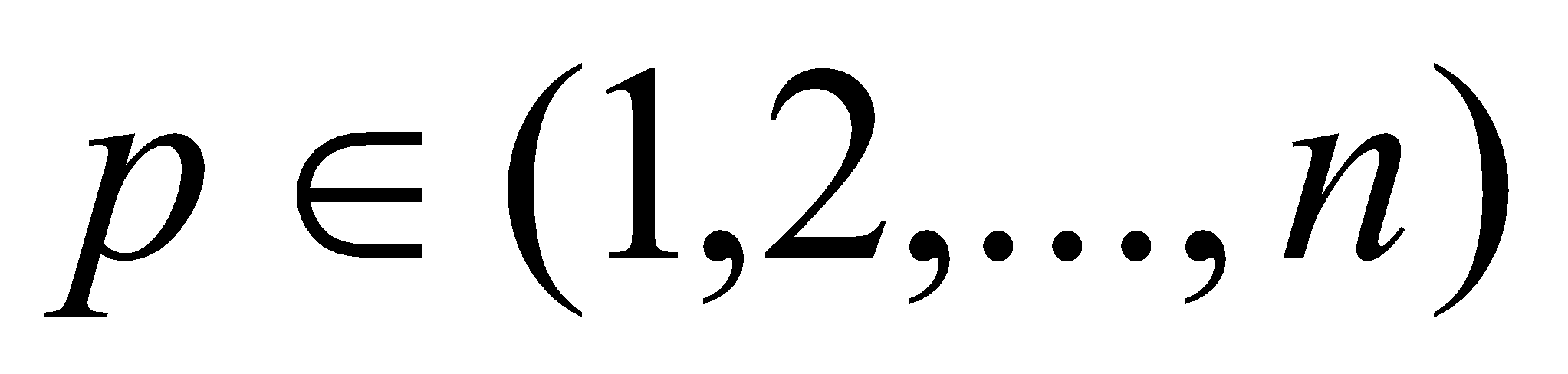
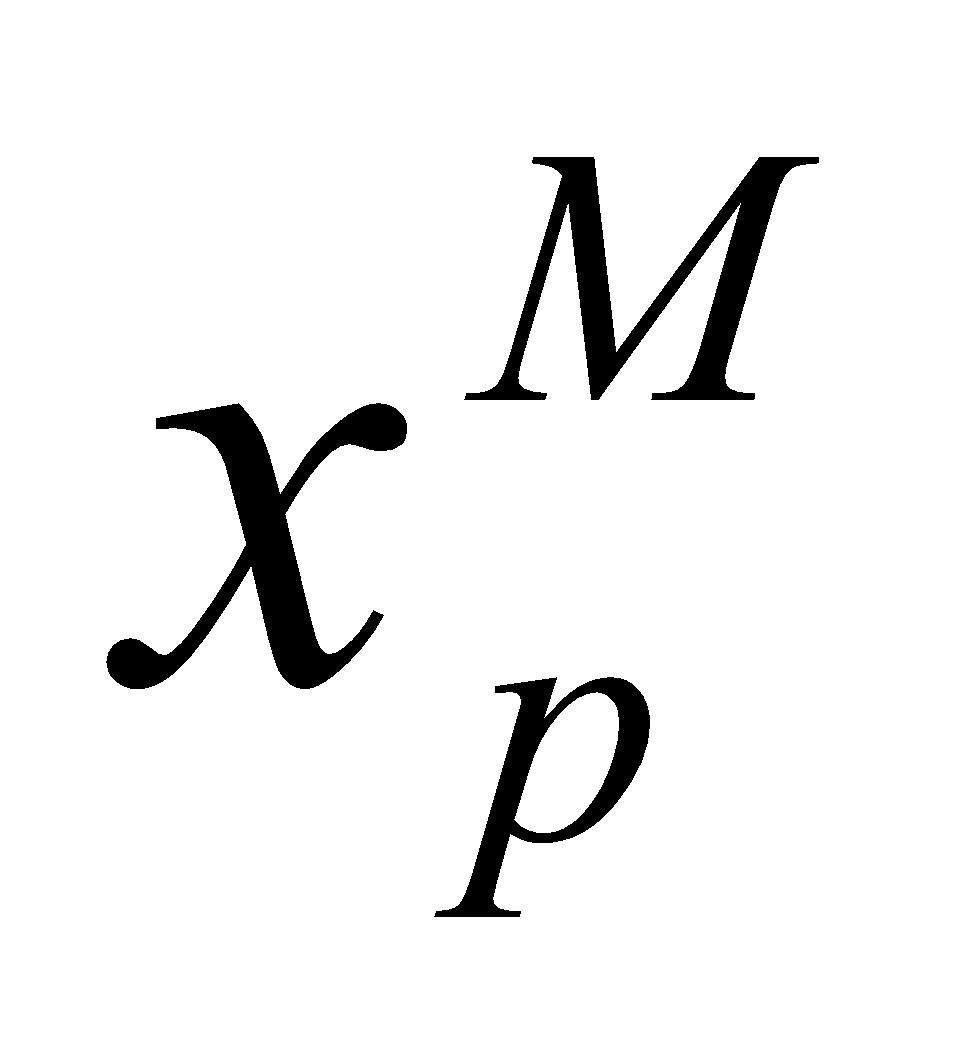
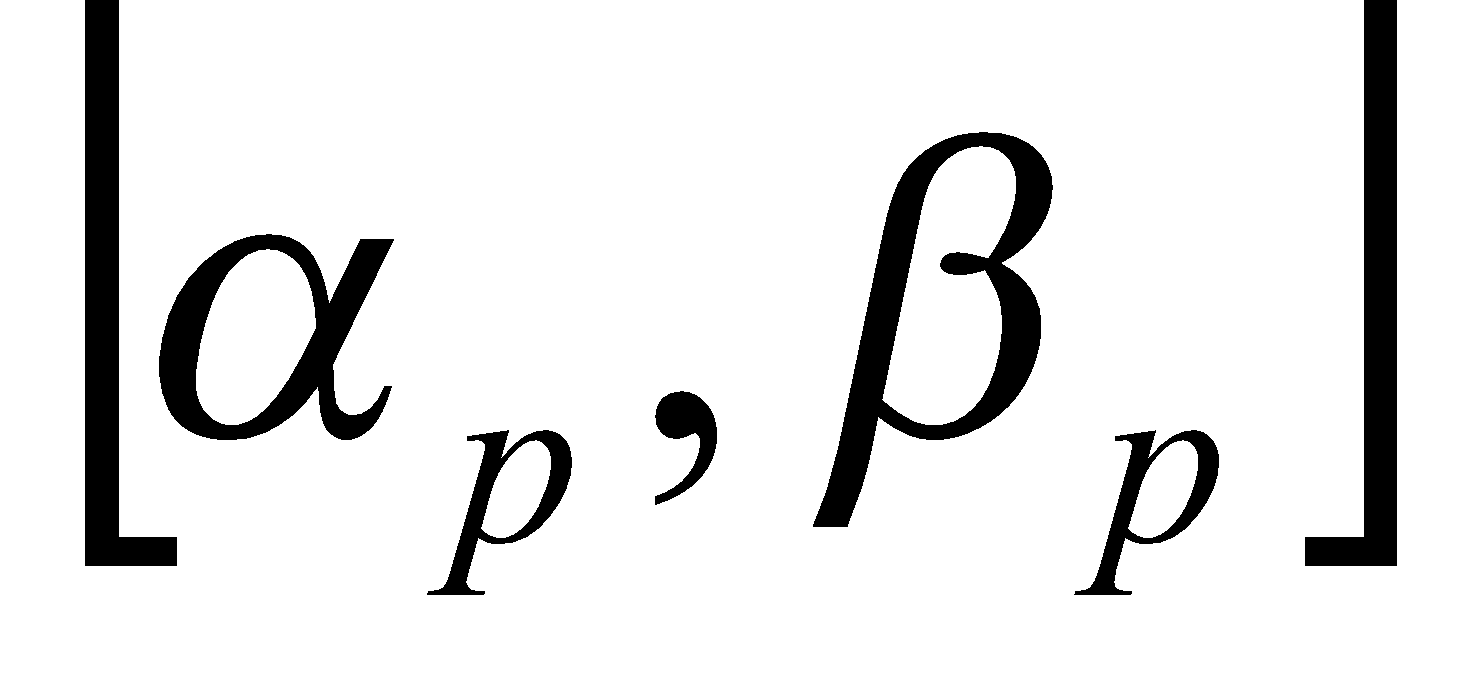
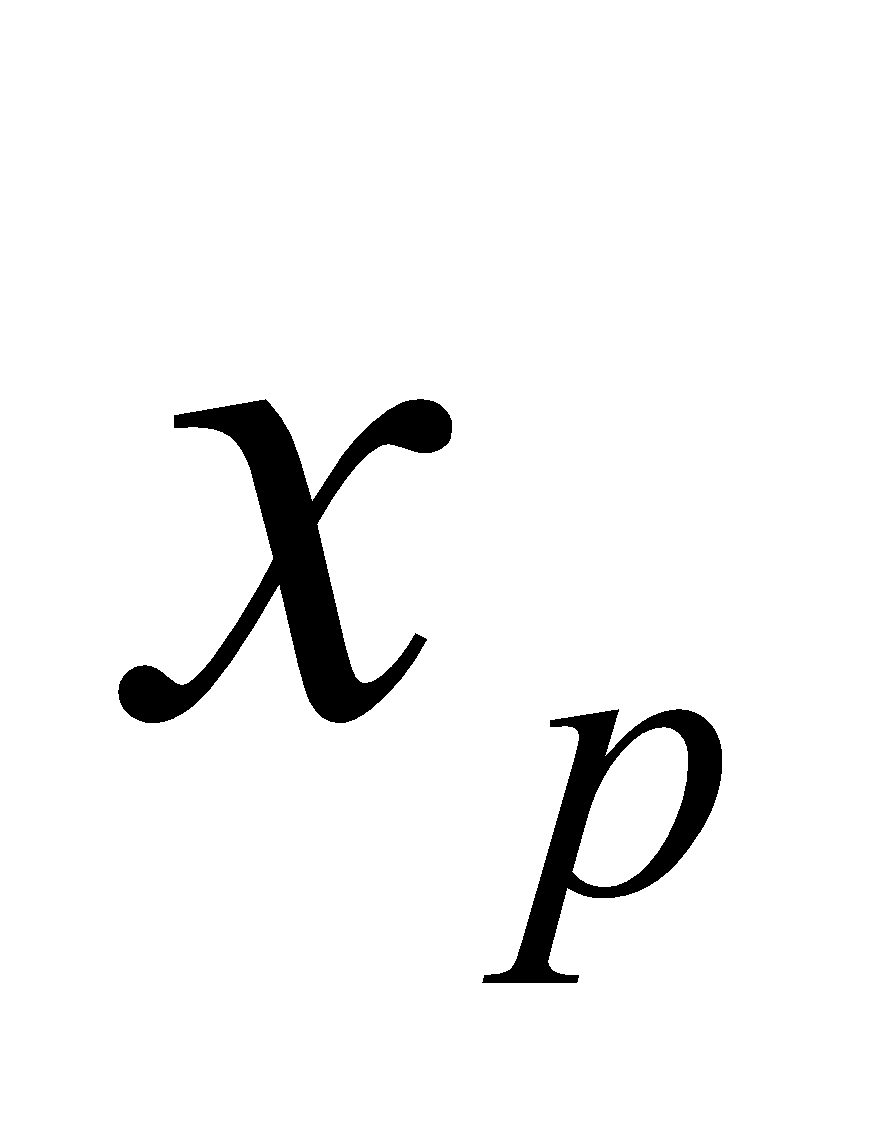
Ш.3.5. Если допустимое множество является выпуклым, то кроссинговер обеспечивает допустимость обоих потомков, в случае если допустимы оба родителя. Следует проверить допустимость каждого потомка перед тем, как он будет включен в новую популяцию. Если оба потомка являются допустимыми, тогда родители заменяются этими потомками. Если это не так, сохраняется допустимый потомок, если он существует, а затем вновь выполняется оператор кроссинговера с новым значением случайного числа  до тех пор, пока не будут получены два новых допустимых потомка или не будет превышено заданное число циклов. В этом случае осуществляется замена родителей только теми (сохраненными ранее) потомками, которые оказались допустимыми.

**Ш.4. Мутация.**

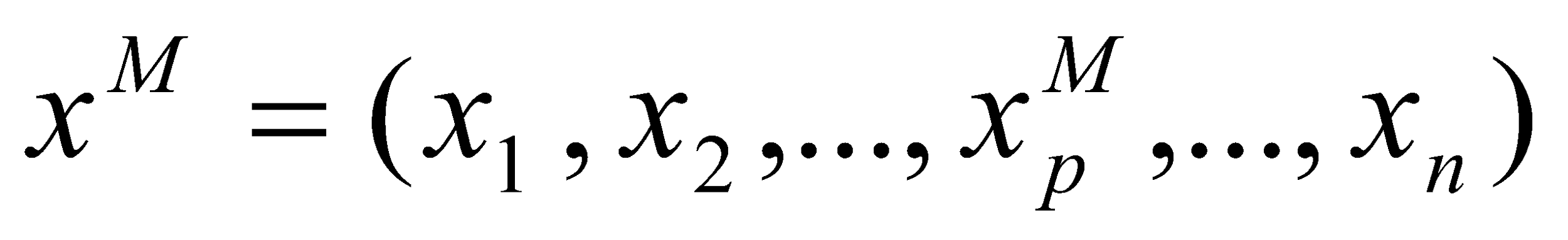
Мутация **–** это преобразование хромосомы , случайно изменяющее один или несколько из её генов. Оператор мутации предназначен для того, чтобы поддерживать разнообразие особей в популяции.

Ш.4.1. Определим параметр  как вероятность мутации. Эта вероятность дает ожидаемое число  хромосом, подвергаемых операции мутации.

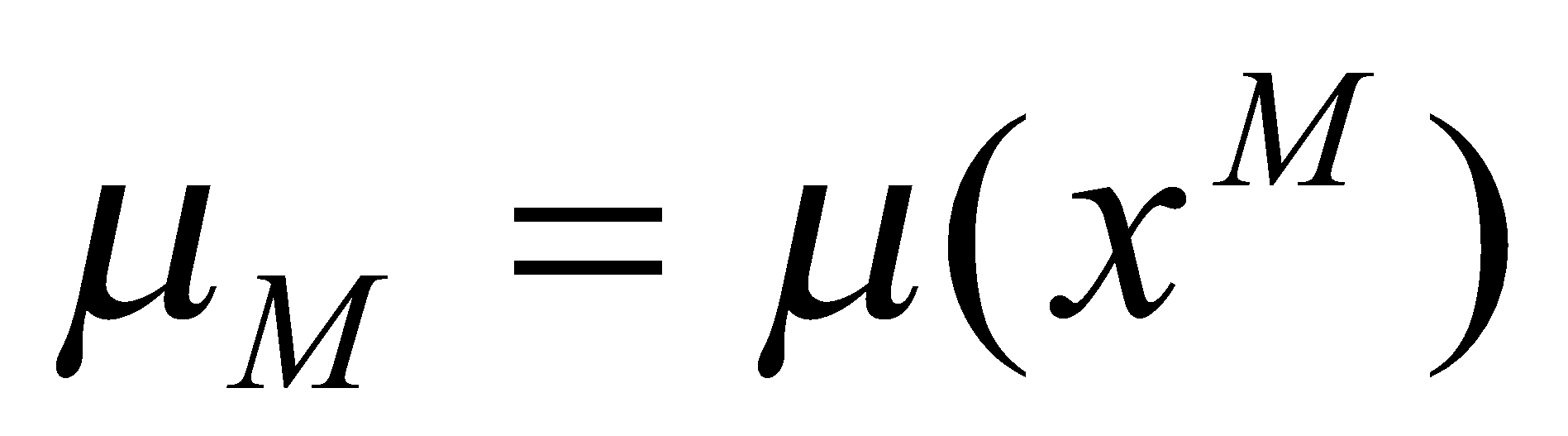
Ш.4.2. Для операции мутации выполняется процесс, повторяющийся от  до : формируется случайное действительное число из сегмента , при этом, если , то хромосома  выбирается как родительская для операции мутации Для каждой выбранной родительской хромосомы , обозначенной как , производится мутация.

Ш.4.3. Поочередно рассматривается каждый потомок из ожидаемого числа  хромосом. Среди генов выбранной родительской хромосомы  случайно ( с вероятностью ) выбирается один с номером  подлежащий замене. Его новое значение  случайным образом выбирается из промежутка  изменения выбранной координаты .

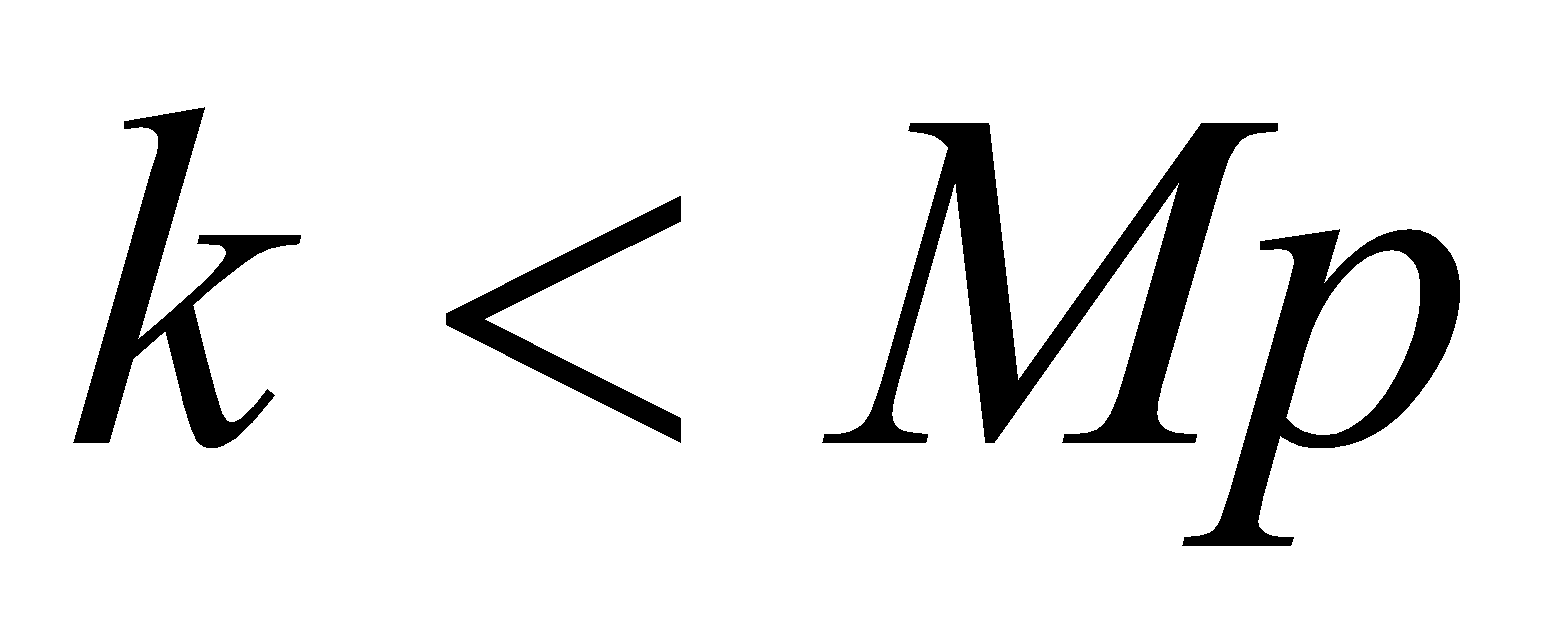
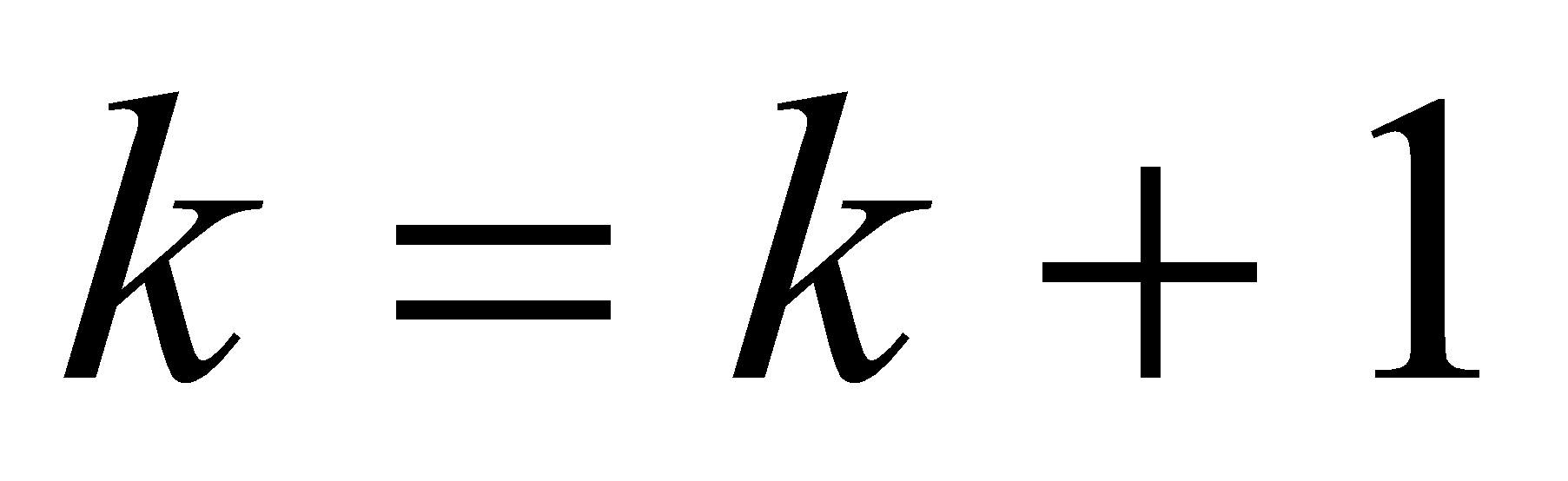
**Ш.5. Формирование новой популяции.**

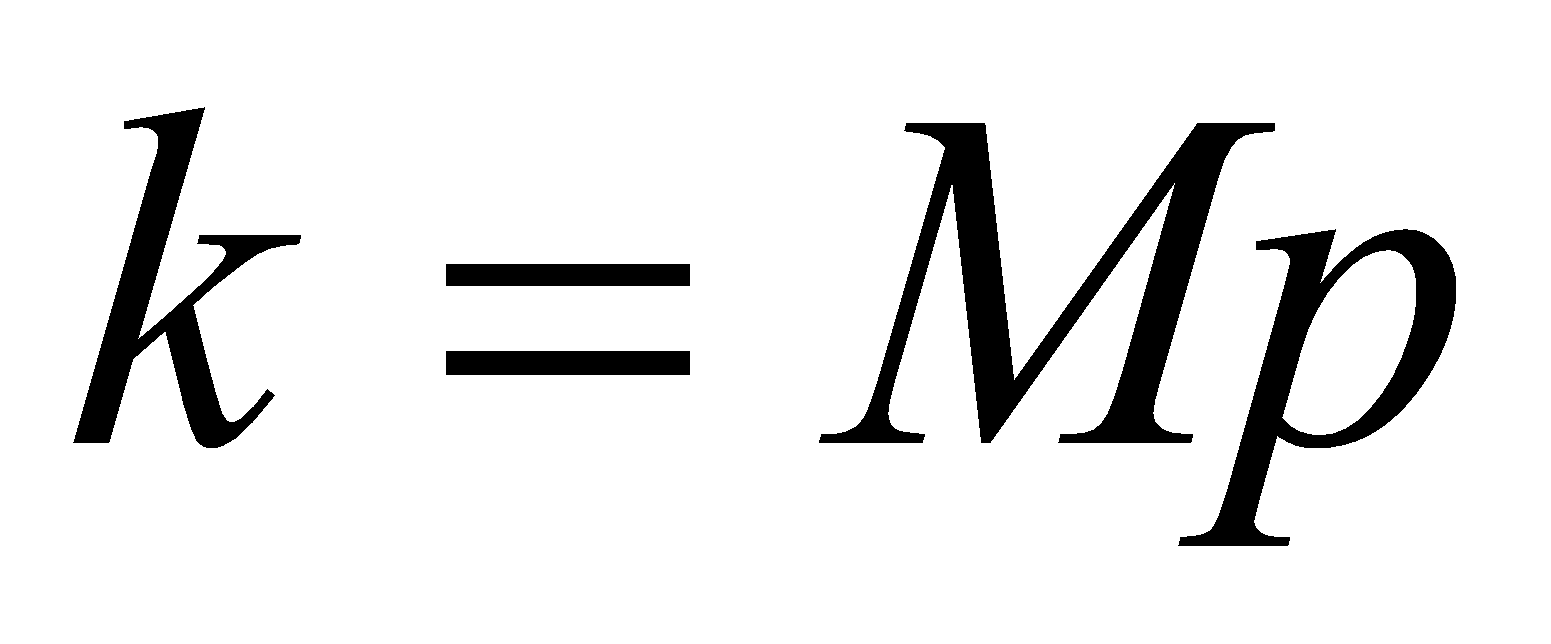
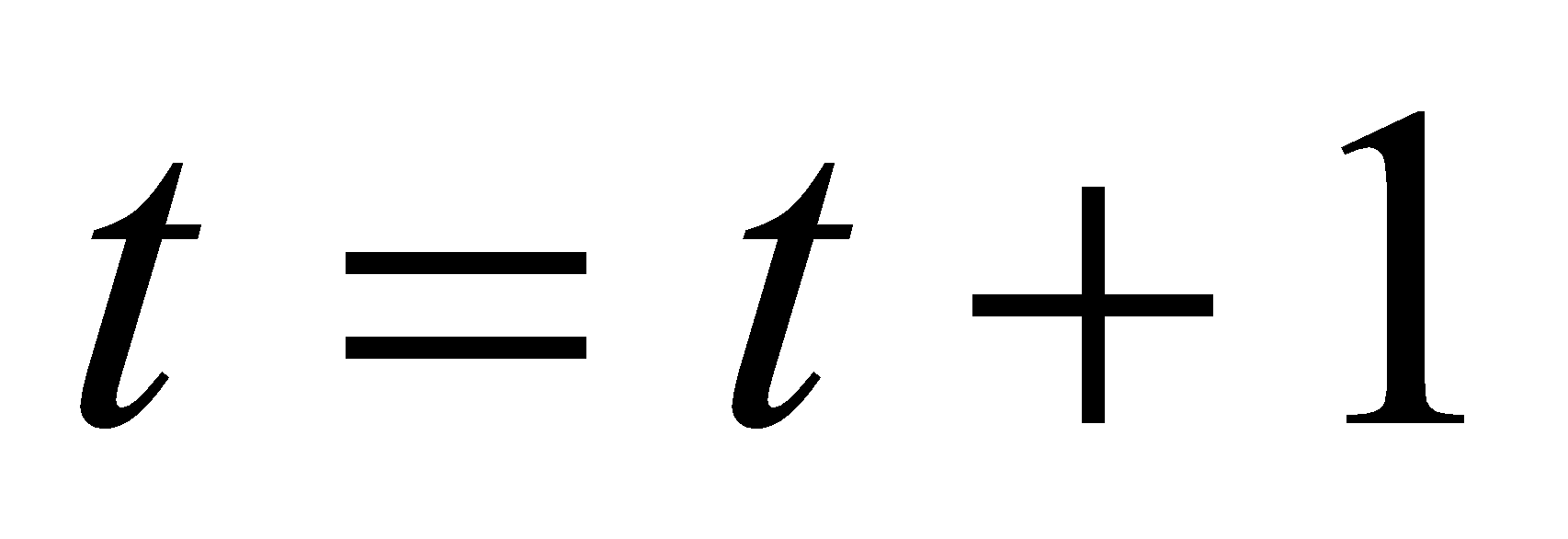
Ш.5.1. С равной вероятностью из потомков мутантов предыдущего шага выбирается один .

Ш.5.2. Выбранный потомок добавляется в популяцию вместо хромосомы, которой соответствует наименьшее значение функции фитнеса (наихудшее из допустимых значений).

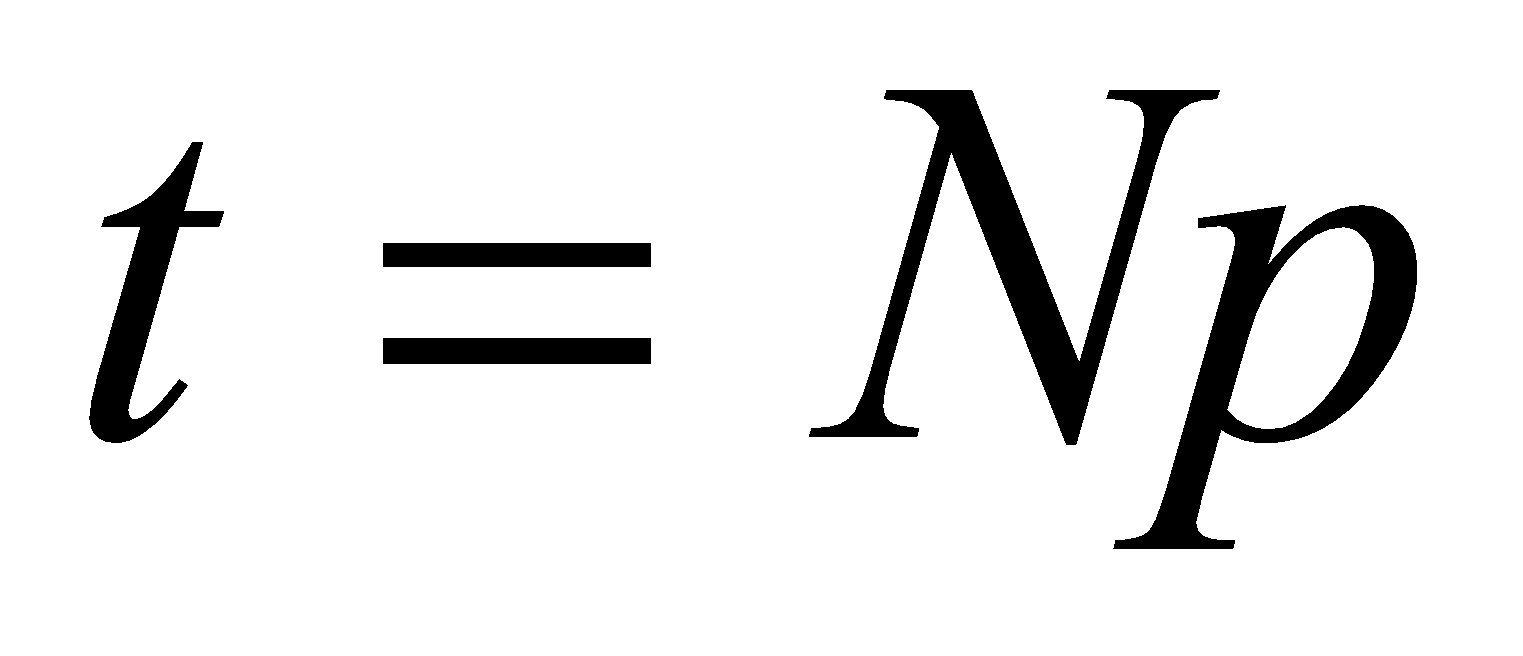
Ш.5.3 Вычисляется значение функции фитнеса для мутантного потомка .

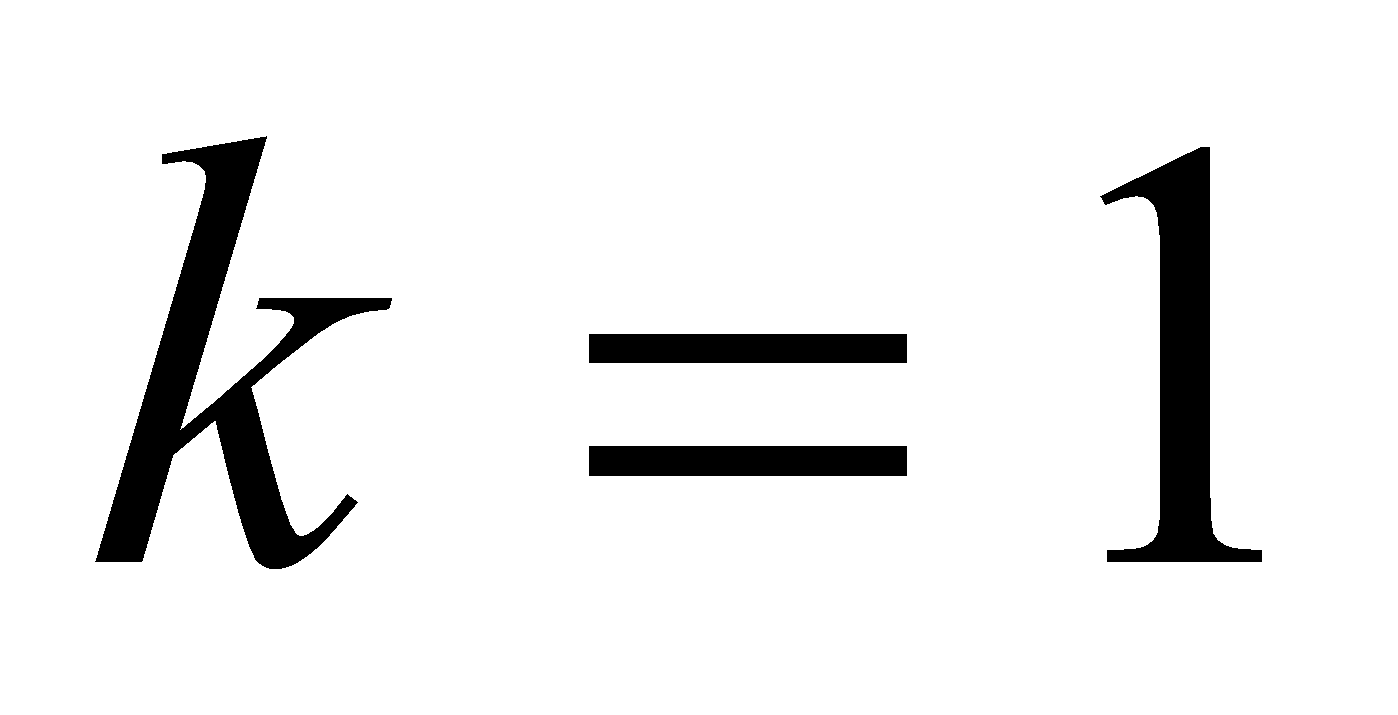
Ш.5.4. Проверка условий:

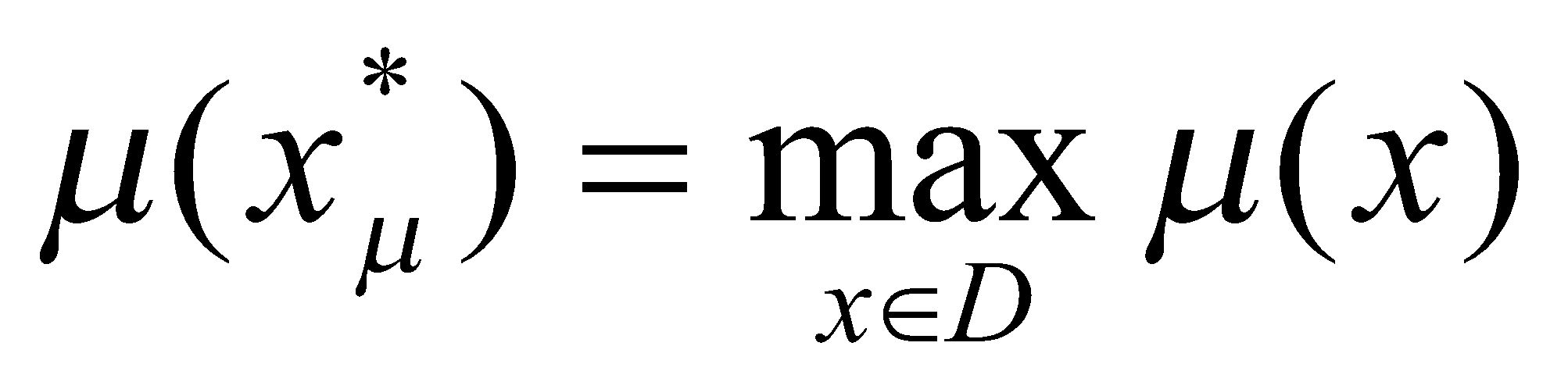
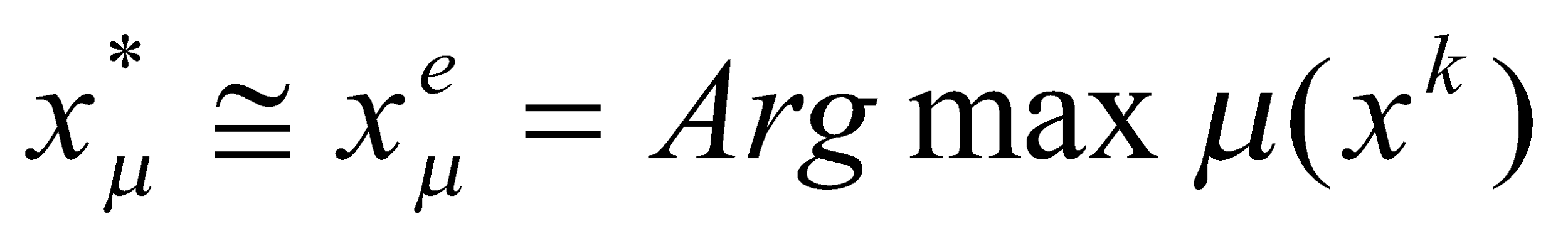
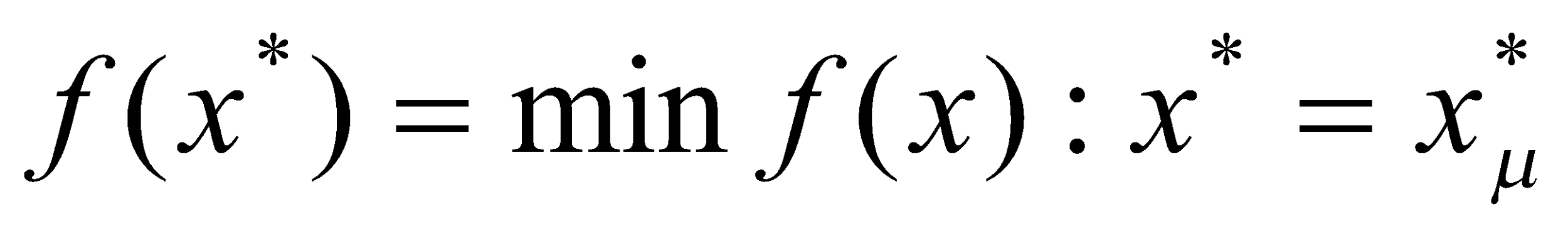
Ш.5.4.1. Если , то  и переход на Ш.2.

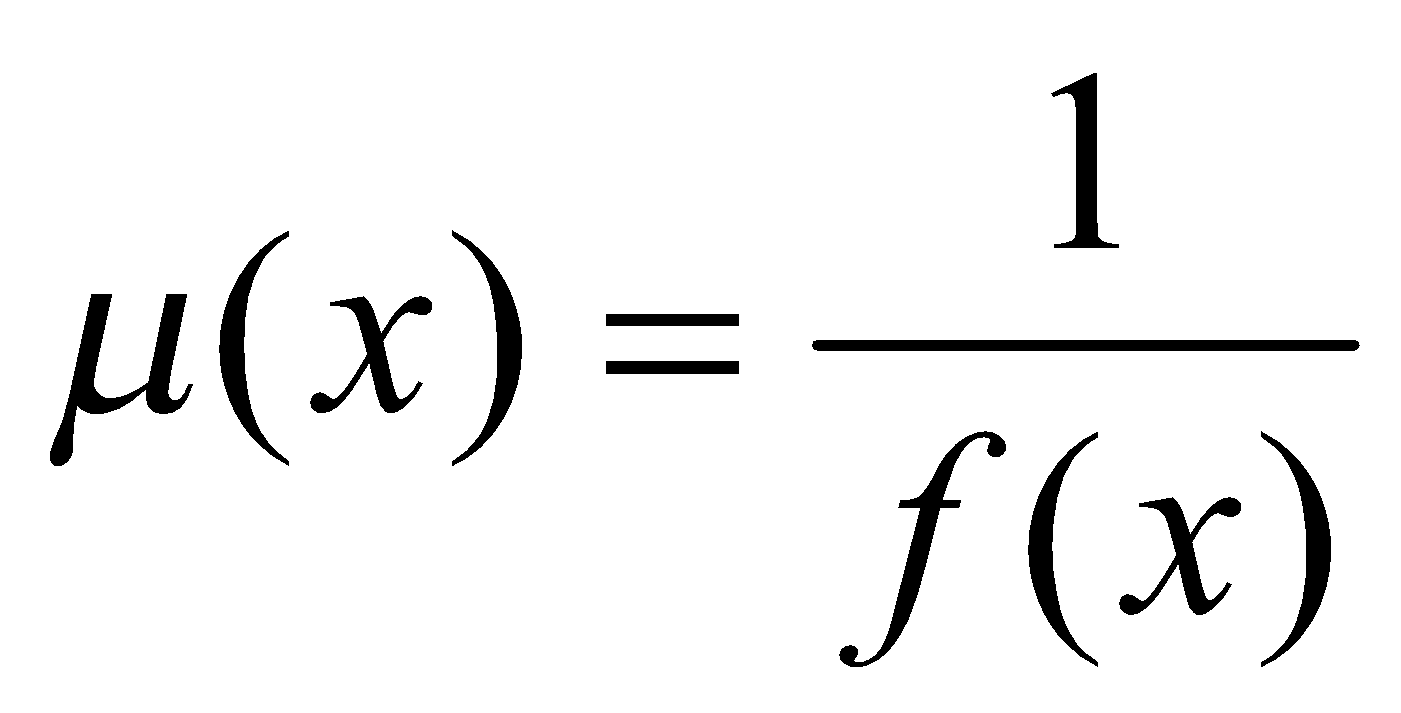
Ш.5.4.2. Если , то  и переход на Ш.6.

**Ш.6. Проверка условия останова генетического алгоритма.**

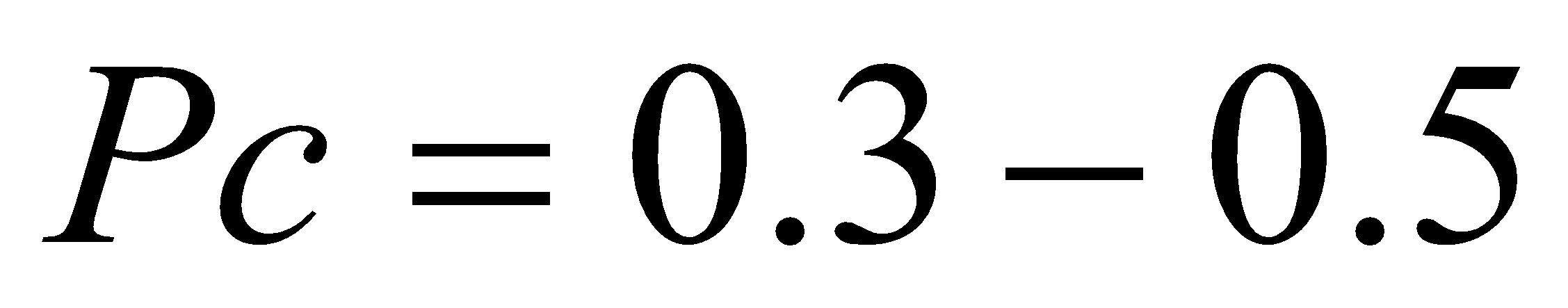
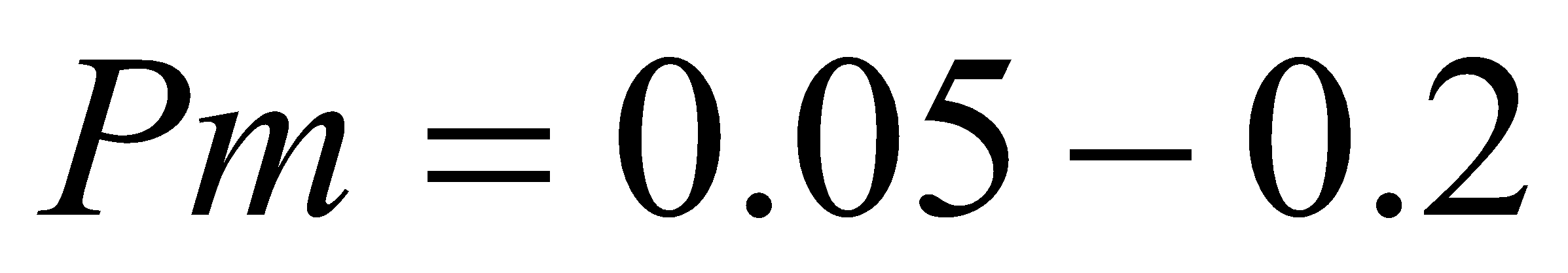
Условием окончания работы генетического алгоритма является формирование заданного количества популяций .

Ш.6.1. Если условие не выполнено, то полагаем  и переход на Ш.2.

Ш.6.2. Если условие окончания работы выполнено, то в качестве решения (приближенного) задачи  выбирается особь с лучшим значением функции фитнеса из текущей популяции:  , а по нему определяется приближенное решение поставленной задачи .

**Замечание 1.** В качестве *функции фитнеса* можно использовать обратную целевую функцию.

**Замечание 2.** Обычно размер популяции выбирают в пределах 30-60 особей.

**Замечание 3.** Вероятность кроссинговера принимается равной , вероятность мутации .