**Задание 3**

*Дано*

Данные хранятся в блоках памяти. Для каждого блока памяти, есть счетчик количества чтений данных из этого блока. При достижении счетчиком порога T, данные из данного блока должны быть скопированы в другой блок. Поскольку блоков много и максимальные значения счетчика для каждого блока очень большие, то требуются значительные ресурсы. Для уменьшения ресурсов может использоваться вероятностный счетчик. Инкремент счетчика происходит не каждый раз при чтении данных из блока, а с некоторой постоянной вероятностью P.

*Найти*

Найти порог для вероятностного счетчика, при достижении которого фактическое количество чтений блока не превысит T с вероятностью 99%.

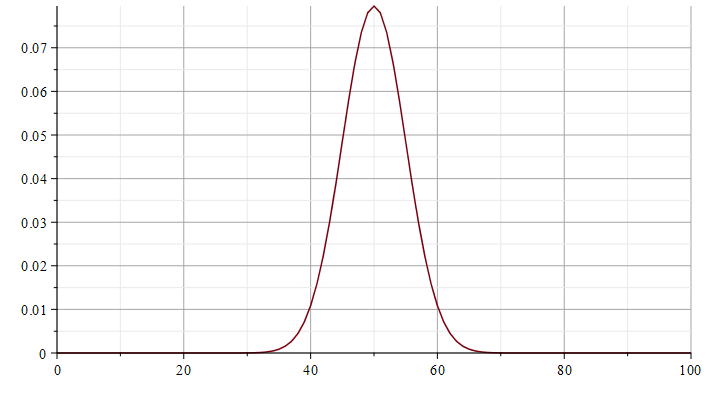
*Решение*

Интерпретируем исходные данные в терминах биноминального распределения. Так пусть реальное число чтений Т - есть число испытаний, а вероятность инкрементации счетчика Р - есть вероятность успеха одного испытания. Всего успешных испытаний S. Тогда распределения вероятности определяется выражением (1):

 (1)

= 

Для наглядности приведем график функции (1) при Т=100 и Р=50%

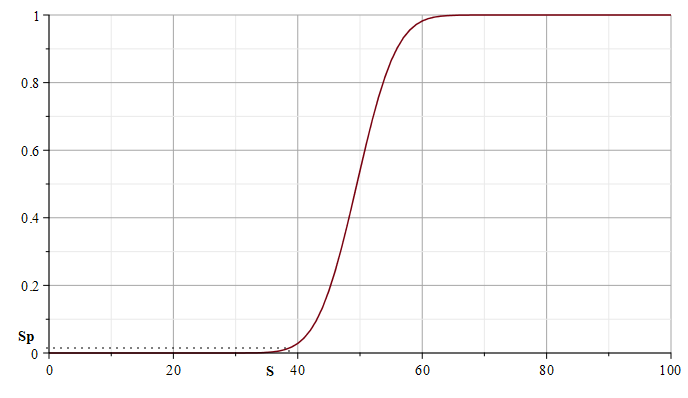


Функция распределения при этом имеет вид (2):

 (2)

 =





Если требуется найти порог S при достижении которго количество испытаний не превысит Т с вероятностью 99%, то мы можем переформулировать данное условие следующим образом. Найти порог S который будет превышен с вероятностью 99 % при проведении Т испытаний, т.е. правый хвост биноминального распределения (1):

 (3)

= 

либо порог S который не будет превышен при проведении Т испытаний с вероятностью 1%, т.е. квантиль уровня 1% (первый процентиль) Sp на функции распределения (2), т.е. порог определенный решением уравнения: 

например для параметров T=100 и P=50% значение S=38

= 0.006016487863

= 0.01048936784

Однако это частные решения, определенные подбором значений - общее решение может быть найдено исходя из того, что при больших количествах испытаний (наш случай) биноминальное распределение можно считать нормальным с мат. ожиданием np и дисперсией np(1-р). Вернемся к условиям задачи - основной причиной введения вероятностного счетчика является очень большое число чтений (скажем больше 1е5) поэтому и вероятность инкрементации вероятностного счетчика желательно сделать маленькой (скажем меньше 5%), тогда можно применить распределение Пуассона с λ=np:

 (4)

= 

 (5)

= 

Из распределения Пуассона вывести решение для порога S будет как минимум не наглядно – воспользуемся квантильной функцией нормального распределения выраженную через обратную функцию ошибок:

 (6)

Для первого процентиля выражение имеет вид:

= 

Вычислим первый процентиль для выражения (6) при Т=1е4 и Р=1%.



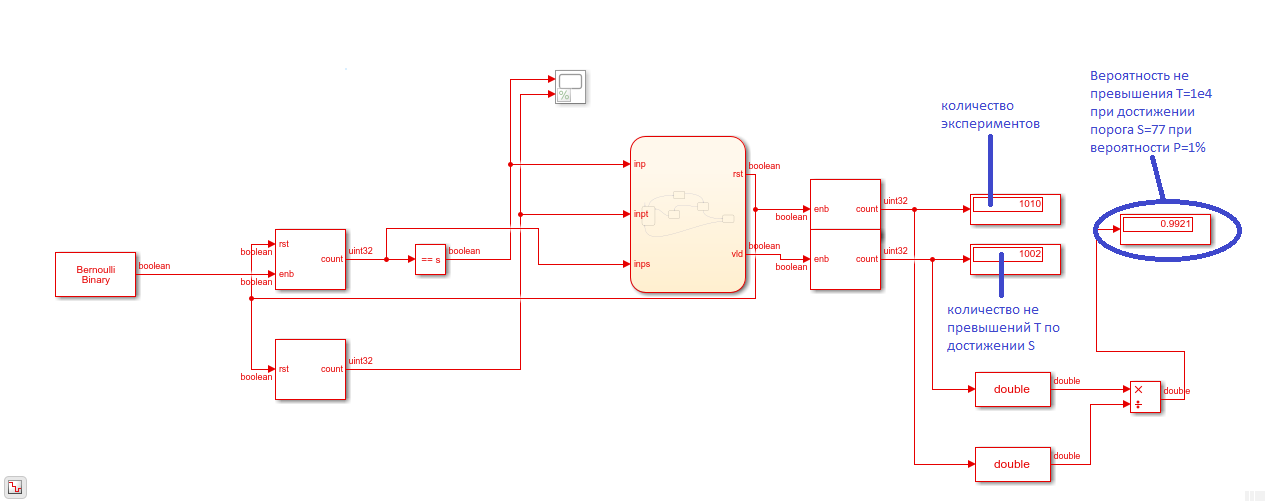
= 77.62272801

Для эксперимента рассчитаем по формулам (2) и (5) вероятность для значений Т=1е4, S=77, Р=1%

= 0.009715096838

= 0.01000799541

и, для подтверждения, смоделируем в MatLab для указаных параметров:



видно что эксперимент подтверждает расчет.