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%.

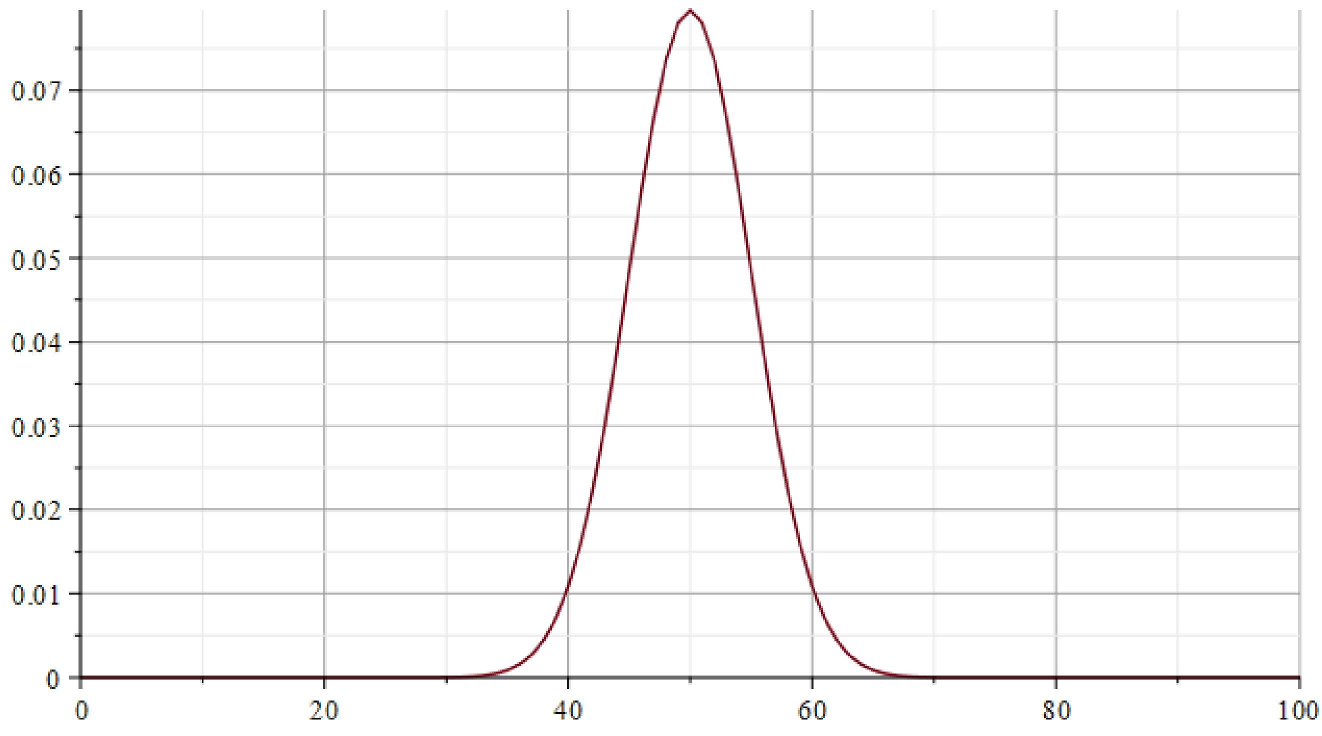
Решение

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

$$pr := (k, n, p) \rightarrow \frac{n!}{(n-k)!\cdot k!}\cdot p^k \cdot (1-p)^{n-k} \tag{1}$$

$$pr(S, T, P) = \frac{T!\, P^S\, (1-P)^{T-S}}{(T-S)!\, S!}$$

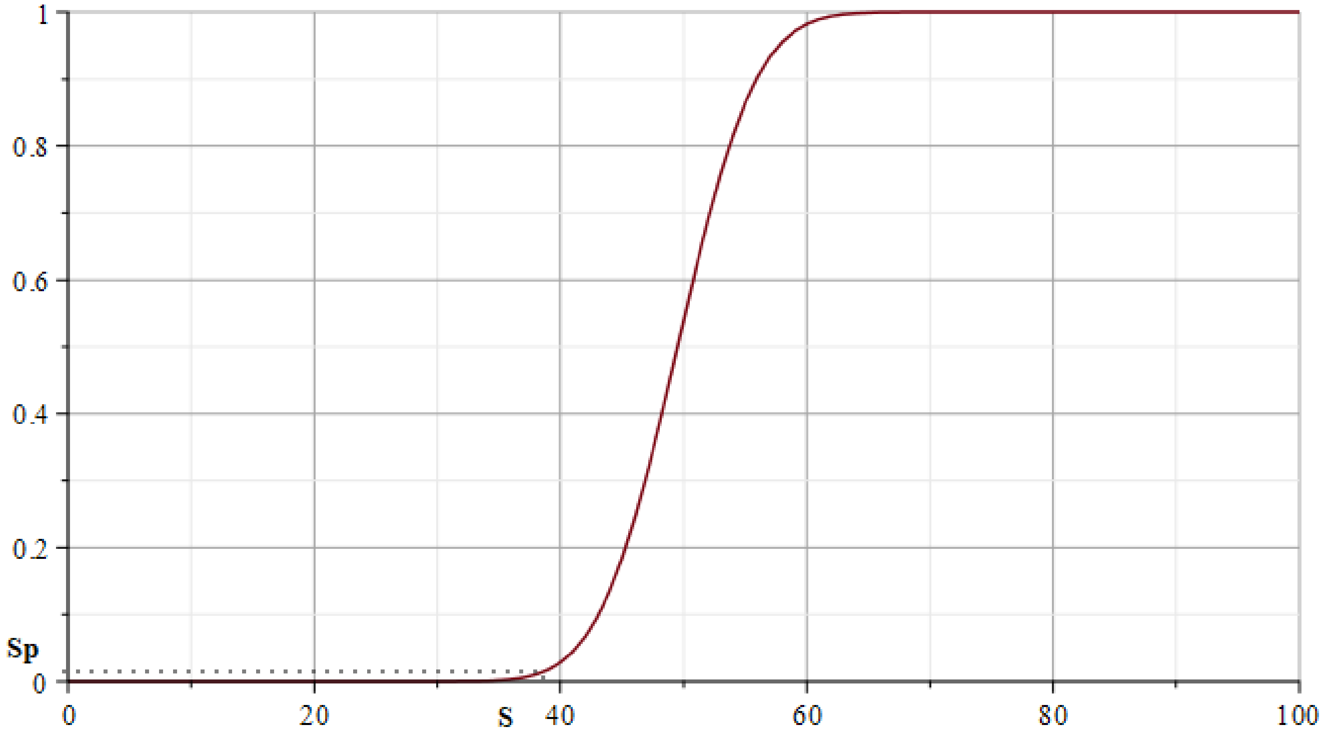
Для наглядности приведем график функции (1) при T=100 и P=50%



Функция распределения при этом имеет вид (2):

$$fr := (k,n,p) \rightarrow \sum_{i=0}^{k} (pr(i,n,p)) \qquad (2)$$

$$fr(S,T,P) = (1-P)^T \left(-\frac{1}{-1+P}\right)^T - \frac{\Gamma(T+1)\, P^{S+1}\, \mathrm{hypergeom}\left([1, -T+S+1], [S+2], \frac{P}{-1+P}\right)}{\Gamma(T-S)\,\Gamma(S+2)\,(1-P)^{-T+S+1}}$$



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

$$Fright := (k,n,p) \rightarrow \sum_{i=k+1}^{n} (pr(i,n,p)) \qquad (3)$$

$$Fright(S,T,P) = \sum_{i=S+1}^{T} \frac{T!\, P^i\, (1-P)^{T-i}}{(T-i)!\, i!}$$

либо порог S который не будет превышен при проведении T испытаний с вероятностью 1%, т.е. квантиль уровня 1% (первый процентиль) Sp на функции распределения (2), т.е. порог определенный решением уравнения: $fr(S,T,P) = 0.01$

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

$fr(37, 100, 0.5) = 0.006016487863$
$fr(38, 100, 0.5) = 0.01048936784$

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-p). Вернемся к условиям задачи - основной причиной введения вероятностного счетчика является очень большое число чтений (скажем больше 1e5) поэтому и вероятность инкрементации вероятностного счетчика желательно сделать маленькой (скажем меньше 5%), тогда можно применить распределение Пуассона с λ=np:

$$prp := (k, \lambda) \rightarrow \frac{\lambda^k}{k!} \cdot e^{-\lambda} \tag{4}$$

$$prp(S, T \cdot P) = \frac{(T P)^S e^{-T P}}{S!}$$

$$frp := (k, n, p) \rightarrow \sum_{i=0}^{k} (prp(i, n \cdot p)) \tag{5}$$

$$frp(S, T, P) = \frac{\Gamma(S+1, T P)}{\Gamma(S+1)}$$

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

$$Finv := (p, \mu, \sigma) \rightarrow \mu + \sigma \cdot \sqrt{2} \cdot (\mathrm{erfi}(2 \cdot p - 1)) \tag{6}$$

Для первого процентиля выражение имеет вид:

$$Finv\left(0.01, T \cdot P, \sqrt{T \cdot P \cdot (1-P)}\right) = T P - 1.590283473 \sqrt{T P (1-P)} \sqrt{2}$$

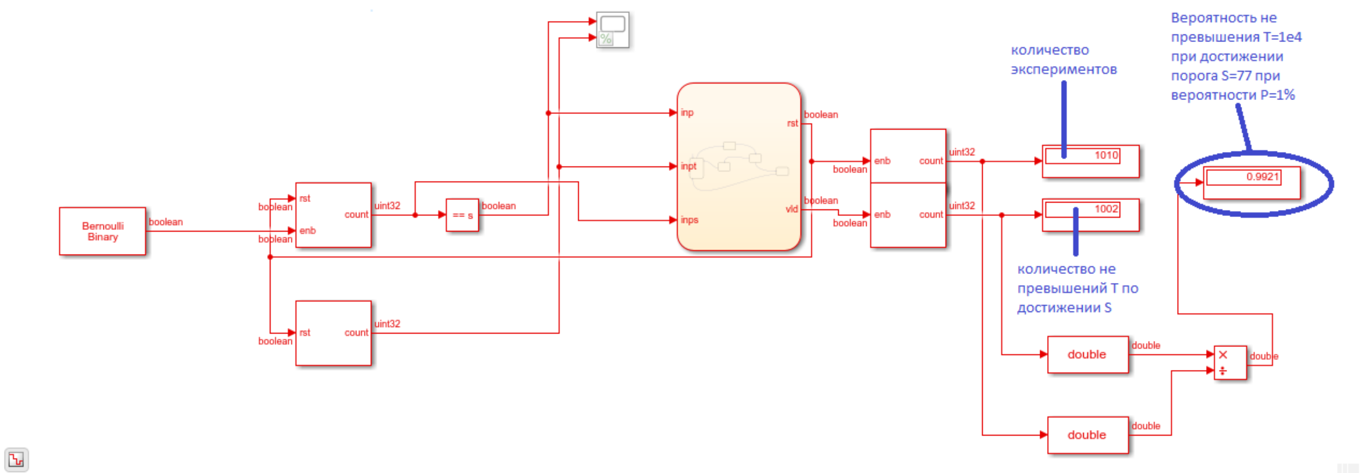
Вычислим первый процентиль для выражения (6) при T=1e4 и P=1%.

$$n := 10000; p := 0.01; q := 1 - p;$$
$$evalf\left(Finv\left(0.01, n \cdot p, \sqrt{n \cdot p \cdot q}\right)\right) = 77.62272801$$

Для эксперимента рассчитаем по формулам (2) и (5) вероятность для значений T=1e4, S=77, P=1%

$fr(77, 10000, 0.01) = 0.009715096838$
$evalf(frp(77, 10000, 0.01)) = 0.01000799541$

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



видно что эксперимент подтверждает расчет.