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. Тогда распределения вероятности определяется выражением **(1)**:

$restart:$

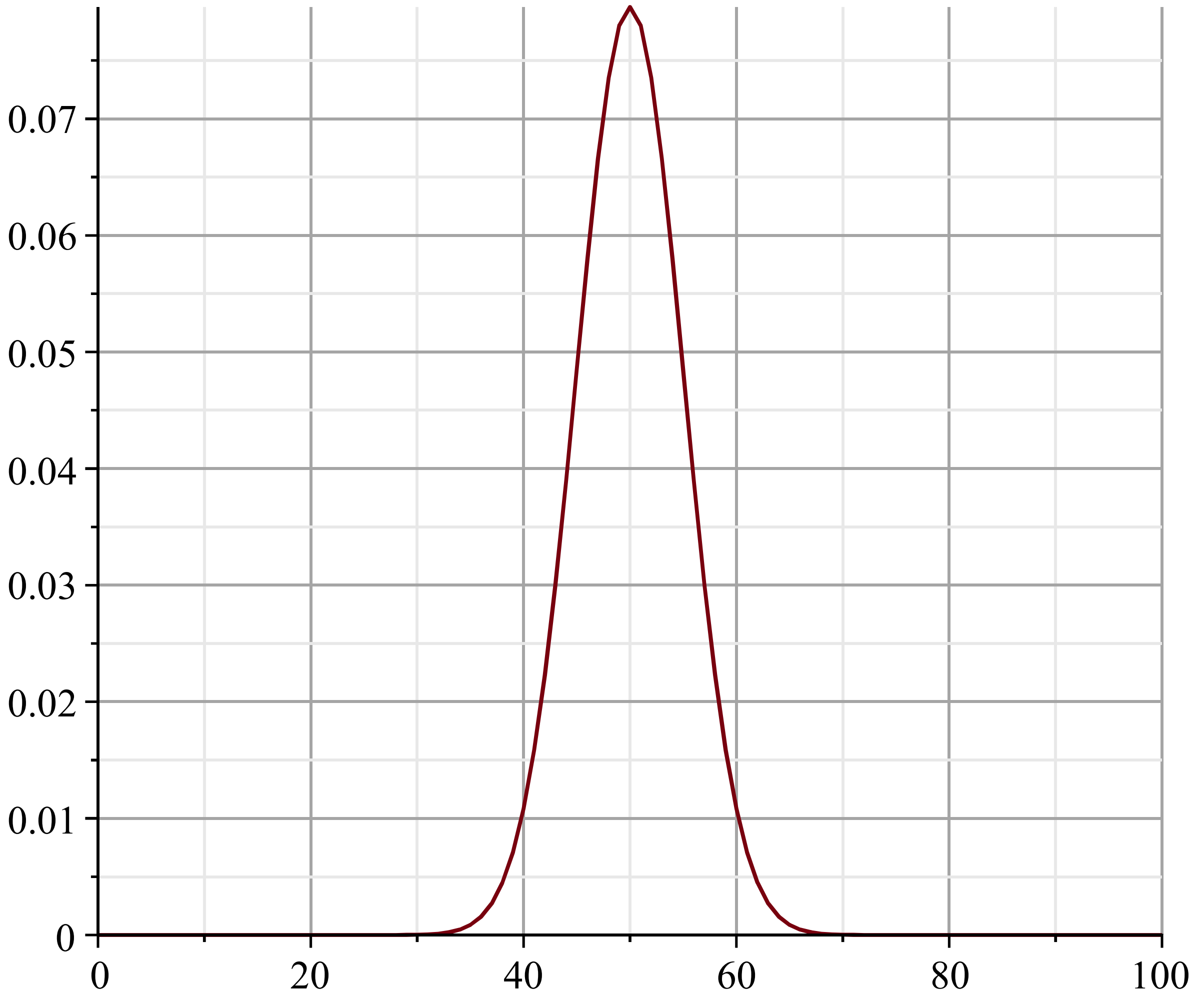
$$pr := (k, n, p) \rightarrow \frac{n!}{(n-k)!\cdot k!} \cdot p^k \cdot (1-p)^{n-k}$$

$$pr := (k, n, p) \mapsto \frac{n!\, p^k\, (1-p)^{n-k}}{(n-k)!\, k!} \quad \textbf{(1)}$$

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

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

$plot(\{seq([x, pr(x, 100, 0.5)], x = 0..100)\})$



Функция распределения при этом имеет вид **(2)**:

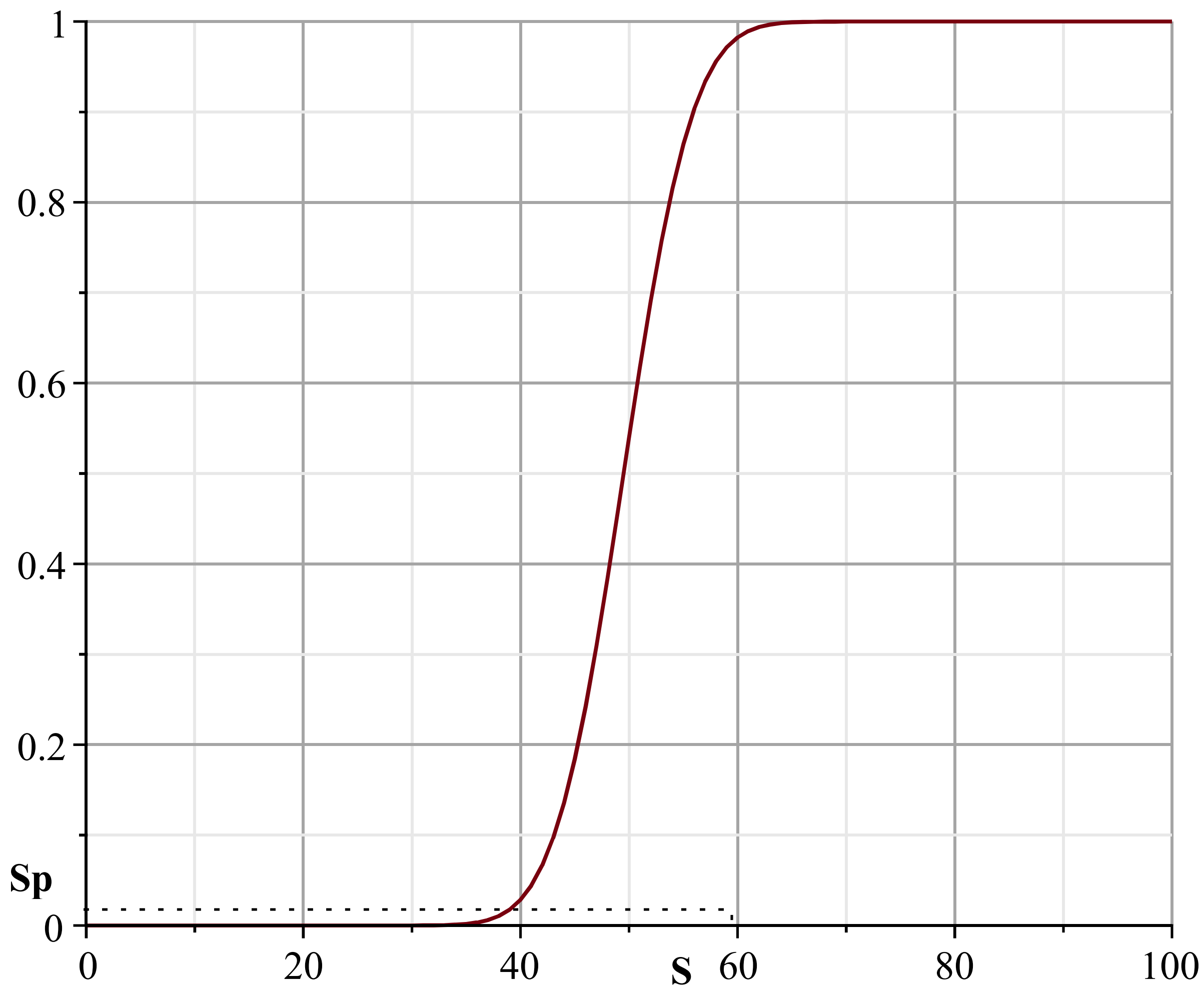
$$fr := (k, n, p) \rightarrow \sum_{i=0}^{k} (pr(i, n, p))$$

$$fr := (k, n, p) \mapsto \sum_{i=0}^{k} pr(i, n, p) \quad \textbf{(2)}$$

$fr(S, T, P) =$

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

$plot(\{seq([x, fr(x, 100, 0.5)], x = 0..100)\})$



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)**:

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

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

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

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:

$fr(S, T, P) = 0.01$

$$(1-P)^T \left(-\frac{1}{-1+P}\right)^T - \frac{\Gamma(T+1)\, P^{S+1} \operatorname{hypergeom}\left([1, -T+S+1], [S+2], \frac{P}{-1+P}\right)}{\Gamma(T-S)\, \Gamma(S+2)\, (1-P)^{-T+S+1}} = 0.01 \tag{4}$$

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

$fr(37, 100, 0.5); fr(38, 100, 0.5)$

$$0.006016487863$$

$$0.01048936784 \tag{5}$$

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). Однако тут также трудно будет найти общее решение. 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:

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

$$prp := (k, \lambda) \mapsto \frac{\lambda^k e^{-\lambda}}{k!} \tag{6}$$

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

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

$$frp := (k, n, p) \mapsto \sum_{i=0}^{k} prp(i, n\, p) \tag{7}$$

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

$s := solve(\mathit{frp}(S, T, P) = 0.01, S)$

$$s := \mathit{RootOf}\left(-100\,\Gamma(\_Z + 1, P\,T) + \Gamma(\_Z + 1)\right) \tag{8}$$