Лекция 1
по Теории Массового Обслуживания

от Татаринова Никиты Алексеевича, БПИ193

2022.10.27

Литература

- D.Cross, C.M.Harris "Fundamentals of Queuing Theory"

Теория массового обслуживания

Родилась из нужд телекоммуникационных компаний – организовать телекоммуникации оптимальным образом. Совмещает теорию вероятностей и исследование операций. 1909г – датский инженер А.К.Эрланг описал модель появления событий во времени – "Theory of Probability and Telephone Conversations".

Что интересует?

- частота звонков (или среднее время между звонками) – интенсивность входящего потока звонков
- среднее время на обслуживание звонка – интенсивность обслуживания
- число телефонисток

Модель

$$\begin{array}{ccccccccc} & & & \nearrow & \boxed{\text{канал } 1} & \searrow & & \\ \text{входящий поток заявок} & \rightarrow & \boxed{\text{очередь}} & \rightarrow & \vdots & \rightarrow & \text{обслуженные заявки} \\ & & \downarrow & \searrow & \boxed{\text{канал } n} & \nearrow & & \\ & & \text{отказ} & & & & & \end{array}$$

Нотация Кендалла

Последовательность из нескольких символов и цифр:

$$A/B/X/Y/Z$$

, который характеризуют некоторые особенности системы массового обслуживания.

- A – закон распределения времени между поступлением заявок
  - M – экспоненциальное (беспамятное распределение)
  - E_k/Er_k – распределение Эрланга порядка k
  - D – детерминированное время
  - G – произвольное распределение (отсутствие ограничений)
- B – закон распределения времени обслуживания (те же обозначения)
- X – число каналов обслуживания (может быть бесконечным, но нетипично)
- Y – ёмкость системы – максимальное число заявок в системе единовременно (может быть бесконечным, но нетипично)

- Z – дисциплина обслуживания – показывает, в каком порядке обслуживаются заявки
  - $FCFS/FIFO$ – First Come First Served
  - $LCFS/LIFO$ – Last Come First Served
  - $RSS/SIRO$ – Random Selection for Service
  - PS – processor sharing (например, 2 канала обслуживают одновременно, но в 2 раза медленнее)
  - PRI – приоритетное обслуживание (заявки разбиваются на разные типы с разными приоритетами)
  - GD – general discipline

Основные случайные величины, связанные с процессом обслуживания

Для числа заявок:

$$\begin{cases} \underset{\substack{\text{число заявок} \\ \text{в системе}}}{N} = \underset{\substack{\text{число заявок} \\ \text{в очереди}}}{N_q} + \underset{\substack{\text{число заявок} \\ \text{на обслуживании}}}{N_s} \\ L = E[N] \quad \text{– среднее число заявок в системе} \\ L_q = E[N_q] \quad \text{– среднее число заявок в очереди} \\ L_s = U = E[N_s] \quad \text{– коэффициент загрузки мощностей} \end{cases}$$

Для времени:

$$\begin{cases} \underset{\substack{\text{время пребывания} \\ \text{заявки в системе}}}{T} = \underset{\substack{\text{время ожидания} \\ \text{в очереди}}}{T_q} + \underset{\substack{\text{время} \\ \text{обслуживания}}}{T_s} \\ W = E[T] \quad \text{– среднее время в системе} \\ W_q = E[T_q] \quad \text{– среднее время ожидания в очереди} \\ \mu = \dfrac{1}{E[N_s]} \quad \text{– интенсивность обслуживания} \end{cases}$$

Кроме того:

$$\begin{cases} \mathbf{P}_{loss} \quad \text{– вероятность потери} \\ \lambda \quad \text{– интенсивность входящего потока заявок} \\ ? \quad \text{– пропускная способность} \end{cases}$$

Пример

Заявки поступают в моменты $0, 20, 40, ...$. Время обслуживания равномерное на отрезке $[14; 24]$ минут. Один канал обслуживания. Если заявка поступает, а канал занят, то она получает отказ. Найти вероятность $P_l(k)$ потери заявки с номером k.

$$\mathbf{P}_l(1) = 0$$

$$\mathbf{P}_l(2) = \frac{24-20}{24-14} = 0.4$$

$$\mathbf{P}_l(3) = \big(1 - P_l(2)\big) \cdot 0.4 = 0.24$$

$$\mathbf{P}_l(4) = \big(1 - P_l(3)\big) \cdot 0.4 = 0.304$$

$$\mathbf{P}_l(k) = \big(1 - P_l(k-1)\big) \cdot 0.4$$

График представляет затухающие колебания вокруг некоего значения. Первые этапы называются переходным периодом, дальнейшие - стационарным периодом.

Геометрическая интерпретация математического ожидания

$$G_T(t) \;=\; \mathbf{P}\big\{T > t\big\} \;=\; 1 - F_T(t)$$ – дополнительная функция распределения (complementary cdf)

Тогда:

$$E\big[T\big] \;=\; \int\limits_0^{+\infty} G_T(t)\cdot dt \;-\; \int\limits_{-\infty}^{0} F_T(t)\cdot dt$$

Для неотрицательной СВ T:

$$E\big[T\big] \;=\; \int\limits_0^{+\infty} G_T(t)\cdot dt$$

Пример 1

Пусть $T \sim Exp(\lambda)$, т.е.

$$G_T(t) \;=\; 1 - F_T(t) \;=\; \begin{cases} 1 & t < 0 \\ e^{-\lambda t} & t \geqslant 0 \end{cases}$$

Тогда:

$$E\big[T\big] \;=\; \int\limits_0^{+\infty} e^{-\lambda t}\cdot dt \;=\; -\frac{1}{\lambda}\cdot e^{-\lambda t}\Big|_0^{+\infty} \;=\; 0 - \left(-\frac{1}{\lambda}\right) \;=\; \frac{1}{\lambda}$$

Пример 2

Пусть:

$$G_T(t) \;=\; 0.7\cdot e^{-t} \qquad t \geqslant 0$$

Найти:

а) долю заявок, которым приходится ждать в очереди

б) среднее время ожидания

в) вероятность, что заявке придётся ждать более 2 минут при условии, что ей придётся ждать

Решение

а)

$$\mathbf{P}\big\{T > 0\big\} \;=\; G_T(0) \;=\; 0.7$$

б)

$$E\big[T\big] \;=\; \int\limits_0^{+\infty} 0.7\cdot e^{-t}\cdot dt \;=\; -0.7\cdot e^{-t}\Big|_0^{+\infty} \;=\; 0 - (-0.7) \;=\; 0.7$$

в)

$$\mathbf{P}\Big(\big\{T > 2\big\}\Big|\big\{T > 0\big\}\Big) \;=\; \frac{\mathbf{P}\Big(\big\{T > 2\big\} \bigcap \big\{T > 0\big\}\Big)}{\mathbf{P}\big\{T > 0\big\}} \;=\; \frac{\mathbf{P}\big\{T > 2\big\}}{\mathbf{P}\big\{T > 0\big\}} \;=\; \frac{G_T(2)}{G_T(0)} \;=\; e^{-2}$$