Санкт-Петербургский государственный политехнический университет

Лабораторная работа № 2

по курсу «Стохастические модели»

«Определение параметров распределения потока заявок
по наблюдениям нагруженности системы»

Студент:	Руцкий В. В.
Группа:	5057/2
Преподаватель:	Иванков А. А.

Санкт-Петербург 2011

1 Постановка задачи

Данной работе производится анализ лога загруженности процессора сервера при поступающих заявках на обработку информации.

В отсутствие заявок величина загруженности процессора представляет собой сумму некоторой постоянной величины загрузки m и случайных отклонений:

$$B(t) = m + \sigma\mathcal{W}(t),$$

где $\mathcal{W}(t)$ — это винеровский процесс.

Интенсивность поступления заявок подчиняются закону распределения Пуассона $\mathcal{P}(\lambda)$.

При поступлении одной заявки нагрузка на процессор мгновенно возрастает, а затем экспоненциально снижается до прежнего уровня. Увеличение загрузки процессора от одной заявки, поступившей в момент времени t_c выражается следующим образом:

$$K_{t_c}(t) = \mathcal{N}(m_c, \sigma_c) \cdot \mathrm{I}(t - t_c) \cdot e^{-\lambda_c(t-t_c)},$$

где $\mathrm{I}(x)$ — фунцкия Хевисайда[1].

В логе загруженности процессора наблюдается общая загрузка процессора:

$$X(t) = B(t) + \sum_{t_c \in T_c} K_{t_c}(t),$$

где T_c — это моменты времени поступления заявок.

Необходимо по дискретным наблюдениям $X(t_i)$

1. оценить моменты времени поступления заявок T_c,
2. оценить параметры модели m, σ, λ, m_c, σ_c, λ_c.

Наблюдения производятся через равные промежутки времени $\Delta t = t_{i+1} - t_i$.

Для упрощения решения λ_c принимается равным величине близкой к нулю, т. е. каждая пришедшая заявка увеличивает загрузку процессора на некоторую фиксированную величину.

[1]

2 Решение

2.1 Оценка моментов времени поступления заявок

Предположим, что в отрезке времени $[t_k, t_{k+n+1}]$ не пришло ни одной заявки. Тогда наблюдения $X(t_k), \ldots, X(t_{k+n+1})$ представляют собой наблюдения $B(t)$. Оценим по этим наблюдениям параметры $B(t)$.

Рассмотрим разности соседних наблюдений — они представляют собой наблюдения нормально распределённой случайной величины:

$$B(t_{i+1}) - B(t_i) = \sigma\mathcal{W}(t_{i+1}) - \sigma\mathcal{W}(t_i) = \sigma\mathcal{N}(0, \Delta t) = \mathcal{N}(0, \sigma\Delta t).$$

Построим точечную оценку $\hat{\sigma}$ методом максимального правдоподобия:

$$\hat{\sigma} = \frac{1}{\Delta t}\frac{1}{n-1}\sum_{i=1}^{n}((X(t_{k+i+1}) - X(t_{k+i})) - 0)^2.$$

Обозначим гипотезу о том, что в промежутке времени $[t_{k+n+1}, t_{k+n+2}]$ не пришло ни одной заявки, как H_0. Тогда

$$\mathcal{L}(X(t_{k+n+2}) - X(t_{k+n+1})|H_0) = \mathcal{N}(0, \sigma\Delta t).$$

В качестве статистики для отвержения гипотезы H_0 возьмём вероятность наблюдения $X(t_{k+n+2})$:

$$T(X_{k+n+2}) = \mathbf{P}\big(B(t_{k+n+2}) = X(t_{k+n+2})\big) = \mathbf{P}\big(\mathcal{N}(0, \hat{\sigma}\Delta t) = X(t_{k+n+2}) - X(t_{k+n+1})\big).$$

[1] Функция Хевисайда: $\mathrm{I}(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geqslant 0 \end{cases}$

А критерием отвержения гипотезы H_0 с уровнем значимости α будет служить следующее выражение:

$$H_0 \text{ rejected} \Longleftrightarrow T(X_{k+n+2}) < \alpha.$$

Алгоритм нахождения моментов времени поступления заявок T_c состоит в следующем:

1. В предположении, что в первые $n+1$ наблюдений не пришло ни одной заявки, оценим $\hat{\sigma}$ и построим критерий для отвержения H_0.

2. Будем добавлять к первым $n+1$ наблюдениям по одному наблюдению и проверять гипотезу H_0. Если H_0 не отвергается, то $\hat{\sigma}$ и критерий для отвержения H_0 пересчитываются для добавленного наблюдения.

3. Как только встретиться наблюдение $n+1+l$, для которого гипотеза H_0 отвергается, то $t_{n+1+l} \in T_c$. Все наблюдения до $t_{n+1+l+1}$ отбрасываются и алгоритм начинается с шага 1 для поиска следующего момента времени прихода заявки.

3 Результаты работы

Список литературы

[1] Г.И. Ивченко and Ю.И. Медведев. *Введение в математическую статистику.* М: Издательство ЛКИ, 2010.