МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ДИНАМИКИ ЭПИДЕМИЧЕСКОГО ПРОЦЕССА

В. В. Евстафьева (Санкт-Петербург, Россия)

Проблема изучения эпидемических процессов, которые ежегодно приносят большой экономический ущерб государству, остается актуальной в настоящее время. Моделирование динамики эпидемического процесса дает возможность прогнозировать уровень заболеваемости на определенный (краткосрочный, среднесрочный или долгосрочный) период времени и тем самым оценить достаточность объема противоэпидемических мероприятий.

Существуют разные математические подходы для исследования поведения такого процесса и для моделирования возникновения, развития и затухания эпидемий используют как системы дифференциальных уравнений, так и эконометрические модели [1]–[4].

В математическую модель, описанную системой дифференциальных уравнений, включается множество различных факторов, от которых зависит уровень заболеваемости, что усложняет возможность оперативного анализа эпидемического процесса.

В настоящей работе демонстрируется эконометрический подход на примере ежемесячных данных по заболеваемости гриппом в 2007 и 2008 годах в Санкт-Петербурге. Данные предоставлены Санкт-Петербургским медицинским информационно-аналитическим центром.

В общем случае в качестве шага (кванта) наблюдения может использоваться неделя или день, что характерно для острой эпидемической ситуации. Важно отметить, что число заболевших в определенный период времени и в определенном регионе – интегрированный показатель, в котором отражены сложные условия и вся совокупность факторов, влияющих на уровень заболеваемости.

Проведен графический и автокорреляционный анализ динамики заболеваемости гриппом. Выявлена сезонность процесса. На данных ряда остатков (после удаления сезонных индексов) построены адаптивные модели Бокса-Дженкинса – авторегрессионные модели проинтегрированного скользящего среднего АРПСС(p, d, q), где p – порядок авторегрессии, d – порядок разности, q – порядок скользящего среднего. Модель выбрана потому, что обладает универсальностью, гибкостью и дает хорошие результаты при краткосрочном прогнозировании.

С помощью автокорреляционного анализа, информационного критерия и оценки точности модели выбраны оптимальные параметры: $p = 2$, $d = 1$, $q = 2$. Модель АРПСС(2, 1, 2), коэффициенты которой рассчитаны по приближенному методу максимального правдоподобия, выглядит следующим образом:

$$\hat{\omega}_\tau = -0,5\omega_{\tau-1} - 0,26\omega_{\tau-2} + \varepsilon_\tau - 0,65\varepsilon_{\tau-1} + 0,98\varepsilon_{\tau-2},$$

где $\tau = \overline{3, n}$, n – количество наблюдений, $\omega_\tau = y_\tau - y_{\tau-1}$, y_τ – уровни исходного ряда без учёта сезонных индексов и среднего значения.

Проверка статистических тестов (в том числе теста по Q-статистике Бокса-Льюнга, теста Харке-Бера) и анализ гистограммы показал, что с вероятностью 95% остатки (после удаления сезонных индексов и значений, рассчитанных по оптимальной адаптивной модели) представляют собой "белый шум".

При моделировании рассматриваемого класса временных рядов (с сезонностью и стохастическим трендом) и при малом количестве наблюдаемых значений принято использовать модели экспоненциального сглаживания. Поэтому для сравнения была построена модель Тейла-Вейджа с оптимальными параметрами, полученными при минимизации средней квадратической ошибки прогноза.

Сравнение краткосрочных прогнозирующих свойств построенных адаптивных моделей подтвердило преимущество моделей Бокса-Дженкинса (4% ошибка прогноза против 16%).

Известно [2], что анализ Фурье благодаря своей простоте удобен для проведения оперативного эпидемического анализа. Аппроксимирующая модель для исходного ряда данных с одной гармоникой выглядит следующим образом:

$$\hat{x}_\tau 298,7 + 3537,48\cos\tau + 4406,59\sin\tau, \quad \tau = \overline{1,n}.$$

Коэффициенты Фурье рассчитаны по методу наименьших квадратов. Построенная модель даже с одной гармоникой хорошо отражает частотные изменения и общую тенденцию наблюдаемого ряда.

Для оценки динамики эпидемического процесса интерес представляет не только аппроксимирующая функция, но и ее первая и вторая производные. Очевидно, что положительный знак перед значением первой производной (скорости) означает, что эпидемия расширяется и в ближайшее время можно ожидать рост числа больных. Отрицательный знак – эпидемия угасает. Положительный знак перед значением второй производной характеризует ускорение прироста числа больных, отрицательный знак – замедление.

В частичной сумме тригонометрического ряда Фурье на основе наименьшей средней ошибки аппроксимации (13,32%) определено оптимальное число гармоник, равное 5. Аппроксимирующая модель с 5 гармониками эффективна для построения долгосрочного прогноза.

Анализ Фурье позволяет создавать математическую модель практически любого эпидемического процесса в определенный момент времени и оценивать, на какой стадии находится процесс. Кроме того, с увеличением числа гармоник, можно с достаточно высокой точностью аппроксимировать этот процесс и прогнозировать его дальнейшее развитие. Однако, модели гармонического анализа уступают моделям Бокса-Дженкинса при построении краткосрочных прогнозов (24% ошибка прогноза против 4%).

Литература

1. Бокс Дж., Дженкинс Г. М. *Временные ряды: Прогноз и управление.* Ч. 1., Ч. 2. М.: Финансы и статистика, 1974.

2. Медик В. А., Токмачев М. С. *Математическая статистика в медицине.* М.: Финансы и статистика, 2007.

3. Славин М. Б. *Методы системного анализа в медицинских исследованиях.* М.: Медицина, 1989.

4. Петунин Ю. И. *Приложение теории случайных процессов в биологии и медицине.* Киев: Наука, 1981.