ОЦЕНКА ПАРАМЕТРОВ МОДЕЛИ INAR(1)

Дорошко О. В., Сталевская С. Н.
Кафедра математического моделирования и анализа данных, Белорусский государственный университет
Минск, Республика Беларусь
E-mail: volha.doroshko@gmail.com

В данной работе приведены методы оценки параметров модели INAR(1). Для оценки самих методов введены понятия смещения и вариации оценки. Представленные в данной статье экспериментальные исследования позволяют получить наглядное представление о свойствах полученных оценок.

Введение

На практике часто встречаются временные ряды, характеристики которых несовместимы с подходом непрерывного моделирования. Целочисленные вариационные временные ряды встречаются во многих контекстах, часто в виде количества событий или отдельных лиц в последовательных интервалах времени. Примерами этого являются число клиентов, ожидающих обслуживания, ежедневное число отсутствующих работников в офисе, число занятых линий телефонной сети, регистрируемое каждый час, число несчастных случаев на производстве каждый месяц, количество сделок с акциями, произведенных за один день и т.д. Для анализа такие рядов в [1] и была введена модель INAR(1).

I. Модель INAR(1)

Используя, введеный в [2], биномиальный оператор прореживания, вместо обычного оператора умножения, используемого в авторегрессионной AR(1), в [1] и [3] была введена неотрицательная целочисленная авторегрессионная модель INAR(1):

$$X_t = \alpha \circ X_{t-1} + \xi_t \qquad (1),$$

где $\alpha \in [0;1]$ и ξ_t - последовательность независимых и одинаково распределенных неотрицательных целочисленных случайных величин с математическим ожиданием $E(\xi_t) = \mu_\xi$ и дисперсией $V(\xi_t) = \sigma_\xi^2$. Оператор прореживания "$\circ$"для X – неотрицательной целочисленной случайной величины и любого $\alpha \in [0;1]$ определяется формулой:

$$\alpha \circ X = \sum_{i=1}^{X} Y_i,$$

где Y_i - последовательность независимых одинаково распределенных случайных величин Бернулли, не зависящие от X, такие, что

$$Pr(Y_i = 1) = 1 - Pr(Y_i = 0) = \alpha.$$

При этом $\alpha \circ X$ имеет биномиальное распределение с условными параметрами $E(\alpha \circ X|X) = \alpha X$ и $V(\alpha \circ X|X) = \alpha(1-\alpha)X$ и безусловными параметрами $E(\alpha \circ X) = \alpha\lambda$, где $E(X) = \lambda$, и $V(\alpha \circ X) = \alpha^2\sigma^2 + \alpha(1-\alpha)\lambda$, где $V(X) = \sigma^2$. Очевидно, что $\alpha \circ X$ принимает только целые значения из [0;X].

II. Оценка параметров модели

1. Оценка Юле-Уокера.
Для выборки размером n модели $X_t, t \in Z$ автокорреляционная функция равна

$$\hat{p}(k) = \frac{\sum_{t=1}^{n-k}(X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^{n}(X_t - \bar{X})^2},$$

где $\bar{X} = \frac{1}{n}\sum_{t=1}^{n} X_t$ – средняя величина. Оценка Юле-Уокера (Yule-Walker estimator(YW)) для α, основывается на том факте, что $p(k) = \alpha^k$:

$$\hat{\alpha} = \hat{p}(1) = \frac{\sum_{t=1}^{n-1}(X_t - \bar{X})(X_{t+1} - \bar{X})}{\sum_{t=1}^{n}(X_t - \bar{X})^2} \quad (1).$$

Первый момент $X_t, t \in Z$ задается, как $E(X_t) = \frac{\mu_\xi}{1-\alpha}$. Используя это, оценку μ_ξ можно получить следующим образом:

$$\hat{\mu}_\xi = (1 - \hat{\alpha})\bar{X},$$

где $\hat{\alpha}$ берется из (1).

2. Условная оценка методом наименьших квадратов.
В качестве альтернативы метода Юла-Уокера можно рассматривать метод оценки условных наименьших квадратов (Conditional least squares estimation (CLS)). Условная оценка наименьших квадратов $\hat{\eta} = (\hat{\alpha}, \hat{\mu}_\xi)^T$ для $\eta = (\alpha, \mu_\xi)^T$ выглядит следующим образом:

$$\hat{\eta} = arg\min_{\eta}(S_n(\eta)),$$

где $S_n(\eta) = \sum_{t=2}^{n}[X_t - g(\eta, X_{t-1})]^2$ и $g(\eta, X_{t-1}) = E(X_t|X_t - 1)$. Таким образом, согласно [4], условная оценка наименьших квадратов для α и μ_ξ может быть представлена, как

$$\hat{\alpha} = \frac{\sum_{t=2}^{n} X_t X_{t-1} - \frac{\sum_{t=2}^{n} X_t \sum_{t=2}^{n} X_{t-1}}{n-1}}{\sum_{t=2}^{n} X_{t-1}^2 - \frac{(\sum_{t=2}^{n} X_t X_{t-1})^2}{n-1}} \quad (2)$$

и

$$\hat{\eta}_\xi = \frac{1}{n-1}\Big(\sum_{t=2}^{n} X_t - \hat{\alpha}\sum_{t=1}^{n} X_{t-1}\Big),$$

где $\hat{\alpha}$ из (2).

3. Оценка максимального правдоподобия. Полагая X_1 фиксированной величиной, условная функция логарифмического правдоподобия выглядит следующим образом:

$$l(\alpha, \mu_\xi) = log\Big(\prod_{t=2}^{n} Pr(X_t|X_{t-1})\Big) = \\ = \sum_{t=2}^{n} log\Big(Pr(X_t|X_{t-1})\Big). \quad (1)$$

Оценки методом максимального правдоподобия (conditional maximum likelihood (CML)) $\hat{\alpha}$ и $\hat{\mu_\xi}$ для α и μ_ξ определяются как значения α и μ_ξ , которые максимизируют условную функцию логарифмического правдоподобия в (3).

III. Смещение и вариация оценок

Для изучения свойств построенных оценок будем рассматривать такие характеристики оценки, как смещение и вариация.
Определение
Смещением статистической оценки $b(\hat{\alpha})$ называется уклонение математического ожидания этой оценки от истинного значения параметра α:

$$b(\hat{\alpha}) = E(\hat{\alpha}) - \alpha.$$

Определение
Вариацией (среднеквадратической ошибкой) статистической оценки $v(\hat{\alpha}$ называется следующее математическое ожидание:

$$v(\hat{\alpha}) = E(||\hat{\alpha} - \alpha||^2).$$

Вариация – всегда скалярная величина.

IV. Экспериментальные исследования

Чтобы понять относительные достоинства каждого из трех методов оценки, были проведены экспериментальные исследования для разных значений параметров $\alpha = 0.1, 0.3, 0.5, 0.7, 0.9$. Значения параметра θ для последовательности μ_ξ приняли равным 1.
CML α была найдена путем максимизации условной функции логарифмического правдоподобия в (3). 200 повторений проводились для каждого из возможных параметров α и были рассчитаны смещение и вариация оценок параметров.
Результаты исследования, представленые в таблицах 1 и 2, показывают соответственно смещение и вариацию при различных значениях α и различных размерах выборки, которые были рассмотрены.

V. Список литературы

1. First–order integer–valued autoregressive (INAR(1)) process / M. A. Al–Osh, A. A. Alzaid // Journal of Time Series Analysis, 1987. –C. 61–275.
2. Discrete analogues of self-decomposability and stability / F. W. Stuetal, K. Van Harn // The Annals of Probability, 1979. –C. 893-899.
3. Some simple models for discrete variate time series / E. McKenzie // Water Resources Bulletin, 1985. – C. 645–650.
4. On conditional least squares estimation for stochastic processes / L. Klimko, P. Nelson // The Annals of Statistic, 1978. C. 629-642.

Таблица 1 – Смещение оценки параметров

		Смещение(α)			Смещение(θ)		
n	α	Y-W	CLS	CML	Y-W	CLS	CML
100	0.1	-0.0457	-0.0339	0.1127	0.096	0.0708	0.0548
	0.3	-0.0852	-0.07	0.1595	0.1788	0.149	0.0747
	0.5	-0.0889	-0.0705	0.1691	0.2534	0.2142	-0.0572
	0.7	-0.0635	-0.0432	0.001	0.2083	0.1527	0.0213
	0.9	-0.0418	-0.0217	-0.0073	0.2886	0.1828	-0.053
200	0.1	-0.0462	-0.0405	0.0957	0.0774	0.0654	0.0357
	0.3	-0.0727	-0.0651	0.2022	0.1707	0.1572	0.029
	0.5	-0.0808	-0.0721	0.035	0.2196	0.201	0.0453
	0.7	-0.0392	-0.0299	-0.0064	0.1269	0.1022	0.042
	0.9	-0.0209	-0.0106	-0.004	0.1603	0.102	-0.0335

Таблица 2 – Вариация оценки параметров

		Вариация(α)			Вариация(θ)		
n	α	Y-W	CLS	CML	Y-W	CLS	CML
100	0.1	0.0108	0.0104	0.1196	0.0264	0.0236	0.0366
	0.3	0.0153	0.0133	0.1636	0.0537	0.0436	0.0867
	0.5	0.0172	0.014	0.0657	0.0993	0.079	0.0957
	0.7	0.0088	0.0066	0.0091	0.0807	0.0601	0.1364
	0.9	0.0041	0.0022	0.0005	0.287	0.1716	0.0334
200	0.1	0.0066	0.0063	0.1577	0.0151	0.0136	0.0253
	0.3	0.0104	0.0094	0.1308	0.0464	0.0415	0.1014
	0.5	0.0107	0.0093	0.0692	0.0648	0.0561	0.0852
	0.7	0.0051	0.0043	0.004	0.0547	0.0456	0.0303
	0.9	0.0012	0.0008	0.0001	0.0997	0.0723	0.0152