УДК 519.2

МНОГОМЕРНАЯ МОДЕЛЬ ФАКТОРА ДОСТОВЕРНОСТИ ДЛЯ ЧИСЛА ИСКОВ[1]

Иванова Н.Л.
Кафедра математической статистики и системного анализа

Поступила в редакцию 20.11.2009, после переработки 09.12.2009.

Рассматривается актуарная эволюционная модель фактора достоверности, где параметром риска является интенсивность поступления исков, зависимым образом меняющаяся во времени. Структура рассматриваемой зависимости порождается сверткой «перекрывающихся» компонент. Исследуются некоторые свойства предложенной модели, в частности, приводится пример слабо стационарной последовательности, определяются параметры оптимального линейного прогноза достоверного среднего, предлагается рекуррентная процедура оценки параметров прогноза.

We offer the credibility model for claim numbers (evolutionary model), where risk parameter is an intensity of a claim amounts, which is allowed to vary by dependent way in successive period. The special type of dependency structure is generated by convolution procedure with «overlapping» components. We investigate some properties of offered model, in particular, the example of weakly stationary sequence is considered, the expression for parameters of the credible mean optimal linear forecast is obtained, the recursive formula for linear forecast parameters is introduced.

Ключевые слова: фактор достоверности, эволюционная модель, многомерное распределение Кокса.
Keywords: credibility, evolutionary cdedibility model, multivariate Cox distribution.

Введение

Одним из вопросов актуарной теории риска является оценка так называемого *достоверного среднего* (credible mean), учитывающего предысторию убыточности конкретного риска при определении величины страховой премии на текущий период. Эта процедура называется оценкой опыта или учетом достоверности (credibility).

Введем основные обозначения. Каждый участник портфеля или коллектива рисков характеризуется *параметром риска* θ. При заданном θ случайная *переменная под риском* ξ имеет распределение $p(x|\theta)$, определенное для $x \in X$ (дискретное или непрерывное).

[1]Работа поддержана РФФИ, проект 08-01-00563.

Для риска с известным параметром θ нетто-премия (или чистая премия) имеет вид

$$m(\theta) = E(\xi|\theta) = \int x\, p(x|\theta)\, dx.$$

Параметр риска имеет плотность $u(\theta)$, $\theta \in \Theta$; предположим, что известны статистики, соответствующие смешанной агрегированной плотности:

$$p(x) = E_\theta(p(x|\theta)) = \int p(x|\theta)\, u(\theta)\, d\theta,$$

в частности, агрегированная нетто-премия

$$m = E_\theta(m(\theta)).$$

Центральной проблемой оценки опыта или фактора достоверности является определение нетто-премии для индивидуального риска на основании только агрегированных статистик и *индивидуальной предыстории* за n лет:

$$\underline{x} = \{\xi_t = x_t;\ t = 1, 2, \ldots, n\},$$

а именно, оценка величины

$$E\{\xi_{n+1}|\underline{x}\}.$$

Различным подходам к решению этой проблемы посвящено много работ (см., например, [2], [3]).

1. Эволюционная модель достоверности

Рассмотрим частный случай модели достоверности, где в качестве переменных под риском рассматриваются суммарные количества значимых исков $N_1, \ldots, N_n$, $\ldots$, соответствующих периодам с номерами $1, 2, \ldots, n, \ldots$, которая называется *эволюционной* моделью достоверности (см. [1], [6]).

Введем специальную форму зависимости во времени параметров риска (интенсивностей заявок) Λ_i, $i = 1, \ldots, n$, соответствующих периодам $1, 2, \ldots, n$. Ее структура описывается многомерным индексом (см. [4], [5])

$$\mathtt{i}^n = (i_1, \ldots, i_n),$$

с помощью которого будем описывать процесс *развития убытков.*

Компоненты i_k, $k = 1, \ldots, n$ индекса $\mathtt{i}^n$ принимают два значения: 1, если иск появился в k-й период, и 0 в противном случае. Обозначим I^n множество возможных значений n-мерного индекса $\mathtt{i}$,

$$I_k = \{\mathtt{i}^n \in I^n:\ i_k = 1\}, \quad I_{kl} = \{\mathtt{i}^n \in I^n:\ i_k = i_l = 1\}$$
$$1_k = \{\mathtt{i}^n \in I^n:\ i_k = 1,\ i_j = 0, \quad j = 1, \ldots, n,\ j \neq k\},$$
$$1_{kl} = \{\mathtt{i}^n \in I^n:\ i_k = i_l = 1,\ i_j = 0, \quad j = 1, \ldots, n,\ j \neq k,\ j \neq l\},$$
$$|\mathtt{i}^n| = i_1 + \cdots + i_n.$$

Считающий вектор, задающий количество значимых исков в каждый из периодов $1, 2, \ldots, n$, определяется набором одномерных случайных переменных $N^{(\mathtt{i})}$, где индекс $\mathtt{i}^n$ идентифицирует периоды с номером k возникновения иска: $i_k = 1$.

Пусть

1) для различных $\mathtt{i}^n = \mathtt{i}$ (будем опускать верхний индекс n, когда ясно, о предыстории какой длины идет речь) переменные $N^{(\mathtt{i})}$ условно независмы при заданных параметрах риска $\{\Lambda^{(\mathtt{i})}\}$;

2) $\Lambda^{(\mathtt{i})}$ независимы и имеют функции распределения $U^{(\mathtt{i})}(\lambda)$;

3) условное распределение величин $N^{(\mathtt{i})}$ при заданных $\Lambda^{(\mathtt{i})} = \lambda$ — Пуассоновское с параметром λ;

4) считающий n-мерный вектор $\mathcal{N}_n$ имеет структуру (см. [4], [5])

$$\mathcal{N}_n = (N_1, \ldots, N_n) = \left(\sum_{\mathtt{i} \in I_1} N^{(\mathtt{i})}, \ldots, \sum_{\mathtt{i} \in I_n} N^{(\mathtt{i})} \right). \tag{1}$$

Тогда распределение j-й ($j = 1, \ldots, n$) компоненты вектора $\mathcal{N}_n$ (имеющего многомерное распределение Кокса) может вычисляться согласно

$$P(N_j = k) = \underbrace{\int_0^\infty \cdots \int_0^\infty}_{I_j} \sum_{\substack{k^{(\mathtt{i})},\, \mathtt{i} \in I_j,\, |\mathtt{i}|>1 \\ \sum_{\mathtt{i} \in I_j} k^{(\mathtt{i})} \le k}} P\left(N^{(1_j)} = k - \sum_{\mathtt{i} \in I_j \colon |\mathtt{i}|>1} k^{(\mathtt{i})} \right) \times \\ \prod_{\mathtt{i} \in I_j \colon |\mathtt{i}|>1} P(N^{(\mathtt{i})} = k^{(\mathtt{i})}) \prod_{\mathtt{i} \in I_j} dU^{(\mathtt{i})}(\lambda^{(\mathtt{i})}). \tag{2}$$

Производящие функции вероятностей (PGF) и моментов (MGF) для j-й компоненты вектора $\mathcal{N}_n$ имеют соответственно вид:

$$F_{N_j}(r) = \prod_{\mathtt{i} \in I_j} F_{N^{(\mathtt{i})}}(r) = \prod_{\mathtt{i} \in I_j} F_{\Lambda^{(\mathtt{i})}}(e^{r-1}), \tag{3}$$

$$M_{N_j}(r) = \prod_{\mathtt{i} \in I_j} M_{N^{(\mathtt{i})}}(r) = \prod_{\mathtt{i} \in I_j} M_{\Lambda^{(\mathtt{i})}}(e^r - 1). \tag{4}$$

Описанная выше процедура, построенная на свертке, определяет форму *структурной функции* $U_n(\lambda_1, \ldots, \lambda_n)$ — совместной функции распределения вектора $\Lambda = (\Lambda_1, \ldots, \Lambda_n)$.

Определим вектор средних и ковариационную структуру вектора параметров риска с компонентами

$$\Lambda_j = \sum_{\mathtt{i} \in I_j} \Lambda^{(\mathtt{i})}, \quad j = 1, \ldots, n\colon$$

$$E(\Lambda_j) = \sum_{\mathtt{i} \in I_j} m^{(\mathtt{i})} := m_j; \quad \mathrm{Var}(\Lambda_j) = \sum_{\mathtt{i} \in I_j} \sigma^{(\mathtt{i})} := \sigma_j = \sigma_{jj};$$

$$\mathrm{Cov}(\Lambda_j, \Lambda_k) = \sum_{\mathtt{i} \in I_{jk}} \sigma^{(\mathtt{i})} := \sigma_{jk},$$

где

$$m^{(\mathtt{i})} = E(\Lambda^{(\mathtt{i})}), \quad \sigma^{(\mathtt{i})} = \mathrm{Var}(\Lambda^{(\mathtt{i})}), \quad j, k = 1, \ldots, n.$$

Теперь вектор средних и ковариационная матрица вектора переменных под риском N_j переписываются в терминах числовых характеристик величин Λ_i и $\Lambda^{(\mathtt{i})}$ следующим образом:

$$E(N_j) = m_j; \quad \mathrm{Var}(N_j) = \sigma_j + m_j; \quad \mathrm{Cov}(N_j, N_k) = \sum_{\mathtt{i} \in I_{jk}} (\sigma^{(\mathtt{i})} + m^{(\mathtt{i})}). \tag{5}$$

Свойство 1. *Линейная оценка количества значимых исков N_{n+1} в период $n+1$ относительно предыстории $N_1, \ldots, N_n$*

$$f_n(N_1, \ldots, N_n) = a_0(n) + \sum_{j=1}^{n} a_j(n) N_j, \tag{6}$$

минимизирующая выражение

$$E(N_{n+1} - f_n(N_1, \ldots, N_n))^2, \tag{7}$$

является решением следующей системы уравнений:

$$a_0(n) = E(N_{n+1}) - \sum_{j=1}^{n} a_j(n) E(N_j) = m_{n+1} - \sum_{j=1}^{n} a_j(n) m_j$$

$$\sum_{j=1}^{n} \mathrm{Cov}(N_i, N_j) a_j(n) = \mathrm{Cov}(N_i, N_{n+1}), \quad i = \overline{1, n}. \tag{8}$$

2. Пример слабой стационарности априорной последовательности

Рассмотрим ситуацию для подмножества A^n множества I^n:

$$A^n = \{\mathtt{i}^n,\ \mathtt{i}^n \in I^n :\ |\mathtt{i}^n| \le 2\},$$

соответствующую варианту, когда иск может появляться не больше, чем в двух периодах.

Тогда k-я компонента считающего вектора $\mathcal{N}_n = (N_1(n), \ldots, N_n(n))$ может быть записана в виде

$$N_k(n) = N^{(1_k)}(n) + \sum_{j=1, j \ne k, |k-j| \le M}^{n} N^{(1_{kj})}(n),$$

где $\ell = |k - j|$ определяет задержку урегулирования, $\ell \le M$, M — заданная максимально возможная задержка.

Считающие вектора для периодов длины n и $n+1$ связаны соотношением

$$\mathcal{N}_{n+1} = \left(\mathcal{N}_n, \sum_{\mathtt{i}^{n+1} \in A^{n+1}} N^{(\mathtt{i}^{n+1})}(n+1), \right)$$

где

$$N^{(\mathtt{i})}(n) = N^{(1_{kj})}(n) = \begin{cases} N^{(1_k)}(n+1) + N^{(1_{k(n+1)})}(n+1), & k = j, \\ N^{(1_{kj})}(n+1), & k \ne j. \end{cases}$$

Определим **структуру зависимости для векторного параметра риска** Λ. Для этого введем множество интенсивностей для независимых компонент $\{\Lambda^{(\mathtt{i})},\ \mathtt{i} \in A\}$, по отношению к которым применяется свертка.

Если иски проявляются один раз в наблюдаемой последовательности, то:

$$\Lambda^{(1_k)}(n) = \Lambda(0) + \Lambda^-(n,k,M) + \Lambda^+(n,k,M),$$

где $\Lambda(0)$ — интенсивность исков, которые урегулируются за один раз; $\Lambda^-(n,k,M)$ — интенсивность исков, возникшие в ненаблюдаемом в имеющейся предыстории прошлом

$$\Lambda^-(n,k,M) = \begin{cases} \sum\limits_{\ell=k}^{M} \Lambda(\ell), & k \le M, \\ 0, & \text{иначе}, \end{cases}$$

здесь $\Lambda(\ell)$ — интенсивность исков, имеющих задержку длиной ℓ, будем предполагать, что эта величина зависит только от длительности задержки; интенсивность будущих исков $\Lambda^+(n,k,M)$, т. е. исков, которые проявятся в ненаблюдаемом в имеющейся предыстории будущем:

$$\Lambda^+(n,k,M) = \begin{cases} \sum\limits_{\ell=n-k+1}^{M} \Lambda(\ell), & k+M > n, \\ 0, & \text{иначе}. \end{cases}$$

Иски с индексами 1_{kj}, $k \neq j$ проявляются два раза в наблюдаемой последовательности, их интенсивность определяется согласно

$$\Lambda^{(1_{kj})}(n) = \begin{cases} \Lambda(|k-j|), & |k-j| \le M, \\ 0, & \text{иначе} \end{cases}$$

При введенных таким образом параметрах интенсивности параметр риска не зависит от номера периода k, $k = 1, \dots, n$, так как

$$\Lambda_k(n) = \Lambda^{(1^k)}(n) + \sum_{j=1, j\neq k}^{n} \Lambda^{(1^{kj})}(n) = \sum_{\ell=-M}^{M} \Lambda(\ell).$$

Определим для введенной модели **числовые характеристики параметра риска** — вектор средних и ковариационную матрицу.

Обозначим

$$E(\Lambda(\ell)) = m_\ell, \quad \mathrm{Var}(\Lambda(\ell)) = r_\ell$$

— среднее и дисперсию величины $\Lambda(\ell)$,

$$E(\Lambda_k(n)) = \sum_{\ell=-M}^{M} m_\ell := m, \quad \mathrm{Var}(\Lambda_k(n)) = \sum_{\ell=-M}^{M} r_\ell := r,$$

— среднее и дисперсию компонент зависимого вектора параметров риска;

$$\mathrm{Cov}(\Lambda_k(n), \Lambda_j(n)) = \mathrm{Var}(\Lambda(|k-j|)) = \begin{cases} r_{|k-j|}, & |k-j| \le M, \\ 0, & \text{иначе} \end{cases}$$

— ковариацию компонент вектора Λ.

Тогда моменты компонент N_k считающего вектора выражаются в терминах моментов и составляющих параметров риска следующим образом:

$$E(N_k(n)) = m, \quad \mathrm{Var}(N_k(n)) = m + r,$$

$$\mathrm{Cov}(N_k(n), N_j(n)) = \begin{cases} m_{|k-j|} + r_{|k-j|}, & |k-j| \le M, \\ 0, & \text{иначе}, \end{cases}$$

$$\mathrm{Cov}(N_k(n), N_{n+1}(n+1)) = \begin{cases} m_{n+1-k} + r_{n+1-k}, & n+1-k \le M, \\ 0, & \text{иначе}. \end{cases}$$

3. Оценка параметров линейного прогноза эволюционной модели фактора достоверности

Найдем выражение для оптимального линейного прогноза в приведенной ситуации.

Обозначим

$$\mathbf{a}(n) = (a_1(n), \dots, a_n(n))', \quad \tilde{\mathbf{a}}(n) = (a_n(n), \dots, a_1(n))',$$

$$\mathbf{C}(n) = (\mathrm{Cov}(N_j, N_k))_{j,k=\overline{1,n}} = (c_{kj}), \quad c_{kj} = \begin{cases} m + r, & k = j, \\ m_{k-j} + r_{k-j}, & |k-j| \le M, \\ 0 & \text{иначе}, \end{cases}$$

$$\mathbf{q}(n) = (q_1, \dots, q_n) = \begin{cases} (m_1 + r_1, \dots, m_M + r_M, 0, \dots, 0)', & M < n, \\ (m_1 + r_1, \dots, m_n + r_n)', & \text{иначе}, \end{cases}$$

$$\tilde{\mathbf{q}}(n) = (q_n, \dots, q_1).$$

Свойство 2. *Вектор оптимальных коэффициентов линейного прогноза* (6) *с учетом фактора достоверности имеет вид*

$$\mathbf{a}(n) = \mathbf{C}^{-1}(n)\tilde{\mathbf{q}}(n). \tag{9}$$

Минимальное значение среднеквадратичной ошибки (7) *линейного прогноза N_{n+1} при заданной предыстории $N_1, \dots, N_n$ равно*

$$s(n) = r + m - \mathbf{q}(n)'\tilde{\mathbf{a}}(n).$$

Для случая слабой стационарности введем **рекуррентную схему** для оценки параметров оптимального линейного прогноза эволюционной модели фактора достоверности.

Обозначим

$$k(n) = q_{n+1} - \mathbf{q}(n)'\mathbf{a}(n), \quad n \ge 1.$$

Свойство 3. *Коэффициенты $a_0(n+1)$, $\boldsymbol{a}(n+1)$ линейного прогноза достоверного среднего $f_{n+1}(N_1, \ldots, N_{n+1})$ удовлетворяют следующим рекуррентрым соотношениям:*

$$a_0(n+1) = \left(1 - \frac{k(n)}{s(n)}\right) a_0(n), \quad a_1(n+1) = \frac{k(n)}{s(n)},$$
$$(a_2(n+1), \ldots, a_{n+1}(n+1))' = \boldsymbol{a}(n) - \frac{k(n)}{s(n)}\, \tilde{\boldsymbol{a}}(n),$$

где начальные значения записываются в виде

$$a_1(1) = \frac{m_1 + r_1}{r + m}, \quad a_0(1) = m(1 - a_1(1)).$$

Среднеквадратичная ошибка прогноза определяется согласно

$$s(n+1) = s(n) - \frac{k^2(n)}{s(n)}, \quad n \geq 1; \quad s(1) = r + m - \frac{(r_1 + m_1)^2}{r + m}.$$

Список литературы

[1] Albrecht, P. (1985) An evolutionary model for claim numbers. Astin Bulletin, Vol 15, No 1, 1–17.

[2] Buhlmann, H. (1969) Experience Rating and Credibility. ASTIN Bulletin 5, 157–165.

[3] Jewell, W.S. (1975) Model Variations in Credibility Theory. In Kahn, 193–244.

[4] Ivanova, N.L. and Khokhlov Yu.S. (2001) Reconstruction of multivariate distribution with Poisson components. Mosc. Univ. Comput. Math. Cybern., No. 1, 37–42.

[5] Ivanova, N.L. and Khokhlov Yu.S. (2005) Multivariate Collective Risk Model. Vestnik MGU, Ser. 15. Vyhisl. matem I kibern. 3, 35–43.

[6] Sundt, B. (1983) Finite Credibility Formula in Evolutionary Models. Scandinavian Actuarial Journal, 106–116.