Исходные предположения о математической модели практикума

В данном задании на основе статистического анализа реальных данных о процессе исков в автостраховании типа ОСАГО проверяются гипотезы о стохастической модели процесса исков, исследуется характеристики процесса эволюции пребывания страхователей в классах российской СБМ, изменения во времени средних сборов премии. При анализе свойств СБМ примем следующие предположения.

Предположение A_1. По каждому полису в интервале $[0, t]$ может возникнуть $N(t)$ исков к страховой компании (СК) на возмещение ущерба пострадавшей в ДТП стороне, если виновником ДТП является транспортное средство, застрахованное в СК. При этом считается, что

$$N(t) \stackrel{d}{=} POIS(\Lambda_0 t), \tag{1}$$

где Λ_0 – случайная величина с функцией распределения

$$\mathcal{L}(\Lambda_0) = \mathcal{U}(\lambda; \theta_0), \quad \lambda > 0, \tag{2}$$

или плотностью

$$\mathcal{L}'(\Lambda_0) = u(\lambda; \theta_0), \quad \lambda > 0, \tag{3}$$

где $\theta_0 \in \Theta$, θ_0 – неизвестное порождающее значение параметра *смешивающей* функции распределения $\mathcal{U}$. Параметрическое пространство Θ может быть как конечномерным подмножеством, так и бесконечномерным (т.е. функция распределения $\mathcal{U}$ может быть и непараметрической).

Из соотношения (1) следует, что

$$\mathbf{P}\{N(t) = k | \Lambda_0 = \lambda\} = \frac{(\lambda t)^k e^{-\lambda t}}{k!}, \quad k \in N_+, \tag{4}$$

а безусловная вероятность

$$\mathbf{P}\{N(t) = k\} = \int_0^\infty \frac{(\lambda t)^k e^{-\lambda t}}{k!} d\mathcal{U}(\lambda; \theta_0). \tag{5}$$

Распределение, задаваемое равенством (5) называют *смешанным пуассоновским* распределением и обозначают

$$POIS(\Lambda_0) \underset{\Lambda_0}{\wedge} ABC, \tag{6}$$

если ABC – принятое обозначение для случайной величины Λ_0. ■

Если функция распределения $\mathcal{U}$ сложна, то представлять распределение случайной величины $N(t)$ в форме (5) весьма затруднительно. Альтернативный подход состоит в том, чтобы использовать для этой цели инструментарий производящих функций

$$\pi(z; N(t)) = \mathbf{E}\{z^{N(t)}; \theta_0\}, |z| \le 1, \tag{7}$$

которые взаимно-однозначно определяются соответствующими распределениями.

Предположение A_2. Базой для статистического анализа и возможной модификации СБМ служат статистические данные y об объединенном числе ЗУ-исков и ЗНУ-исков на оплату ущерба, собранные по ЭПО объема n за год функционирования страховой компании в системе СБМ. Всюду дальше $t = 1$ (год), $N(1) = N$. Статистические данные y представлены в виде таблицы частот B, содержащей величины ν_i – число клиентов, породивших за год ровно i исков, $i = 0, 1, 2, \dots, \sum_{i=0}^{\infty} \nu_i = n$, т.е. $y = (\nu_0, \nu_1, \dots, \nu_{k_0})$, где $k_0 = \max\{j \geq 0 : \nu_j > 0\}$. ■

Предположение A_3. Число страхователей в ЭПО равно "n" и год от года не меняется. В то же время оно достаточно велико для возможности привлечения асимптотических методов статистического анализа данных, т.е. $n \to \infty$. ■

Предположение A_4. Относительно структуры стохастической модели M, порождающей частотные данные y таблицы B, выдвинуты следующие гипотезы:

M_1 : данные порождены семейством пуассоновских распределений $\mathcal{POIS}$:

$$\pi(z; POIS(\alpha_0)) = \exp\{-\alpha_0(1-z)\}, \quad \alpha_0 > 0; \tag{8}$$

M_2 : данные порождены семейством отрицательно биномиальных распределений $\mathcal{NBIN}$:

$$\pi(z; HOFM(\alpha_0, \beta_0, 1)) = \pi\left(z; NBIN\left(\frac{\alpha_0}{\beta_0}; \frac{1}{1+\beta_0}\right)\right) = \tag{9}$$

$$= \left(\frac{1}{1+\beta_0(1-z)}\right)^{\frac{\alpha_0}{\beta_0}}, \ \alpha_0 > 0, \quad \beta_0 > 0;$$

M_3 : данные порождены семейством пуассон/ обратно-гауссовских распределений $\mathcal{POIS}/\mathcal{INVG}$:

$$\pi(z; HOFM(\alpha_0, \beta_0, \frac{1}{2})) = \pi\left(z; POIS(\lambda) \underset{\Lambda_0}{\wedge} INVG\left(\frac{2\alpha_0}{\beta_0}; \beta_0\right)\right) =$$

$$= \exp\left\{\frac{2\alpha_0}{\beta_0}\left(1 - \sqrt{1+\beta_0(1-z)}\right)\right\}, \quad \alpha_0 > 0, \quad \beta_0 > 0; \tag{10}$$

M_4 : данные порождены семейством распределений Пойа-Эпли $\mathcal{POLAEP}\left(\frac{\alpha}{\beta}; \frac{\beta}{1+\beta}\right)$;

$$\pi(z; HOFM(\alpha_0, \beta_0, 2)) = \pi\left(z; POLAEP\left(\frac{\alpha_0}{\beta_0}; \frac{\beta_0}{1+\beta_0}\right)\right) =$$

$$= \exp\left\{\frac{\alpha_0}{\beta_0}\left(\frac{1}{1+\beta_0(1-z)} - 1\right)\right\}, \tag{11}$$

$\alpha_0 > 0, \quad \beta_0 > 0;$

M_5 : данные порождены семейством распределений $\mathcal{HOFMAN}$'а общего вида с производящей функцией

$\pi(z; HOFM(\alpha_0, \beta_0, \gamma_0)) = \exp\{-\psi((1-z); \alpha_0, \beta_0, \gamma_0\}$,

где $\alpha_0 > 0, \beta_0 > 0, \gamma_0 \geq 0, \theta_0 = (\alpha_0, \beta_0, \gamma_0)$,

$$\psi(t;\alpha,\beta,\gamma) = \begin{cases} \alpha t \ \text{ при } \gamma = 0, \\ \frac{\alpha}{\beta(1-\gamma)}\left((1+\beta t)^{1-\gamma} - 1\right) \ \text{ при } \gamma \neq 1, \\ \frac{\alpha}{\beta}\ln(1+\beta t) \text{ при } \gamma = 1. \end{cases} \quad (12)$$

M_6 : данные порождены смесью пуассоновского с неизвестным параметром λ_0 и вырожденного в нуле распределения :

$$\pi(z; ZPOIS(\lambda_0,\varepsilon_0) = \varepsilon_0 + (1-\varepsilon_0)\exp\{-\lambda_0(1-z)\}, \quad (13)$$

где $\lambda_0 > 0, \quad 0 < \varepsilon_0 < 1, \quad \theta_0 = (\lambda_0, \varepsilon_0)^T$;

M_7 : данные порождены смесью пуассоновского и непараметрического распределения на R^1_+. Смотри предположение A_1.

M_8 : данные порождены распределением Хофмана с известным параметром $\tilde{\gamma}$:

$\pi(z; HOFM(\alpha_0,\beta_0,\tilde{\gamma})) = \exp\{-\psi((1-z);\alpha_0,\beta_0,\tilde{\gamma}\}$,

где $\alpha_0 > 0, \beta_0 > 0, \tilde{\gamma} \neq 0, \dfrac{1}{2}, 1, 2, \tilde{\gamma}$ – известная величина,

$\theta_0 = (\alpha_0, \beta_0)$;

$$\psi(t;\alpha,\beta,\gamma) = \begin{cases} \alpha t \ \text{ при } \gamma = 0, \\ \frac{\alpha}{\beta(1-\gamma)}\left((1+\beta t)^{1-\gamma} - 1\right) \ \text{ при } \gamma \neq 1, \\ \frac{\alpha}{\beta}\ln(1+\beta t) \text{ при } \gamma = 1. \end{cases} \quad (14)$$

M_9 : данные порождены смесью отрицательно биномиального и обобщенного Парето распределения.

Предположение A_5. Начальная премия страхователя $S_1 = 1$. Премия S_i в i-тый год страхования определяется классом СБМ, к которому будет приписан страхователь, $\mathcal{L}(S_\infty)$ – предельное распределение премии страхователя.

Предположение A_6. Используется СБМ, описанная в разделе 2.2 главы "2.Организационные и финансовые характеристики ОСАГО". Ее характеристики содержатся в таблице 1 и таблице 2.

Набор статистик критерия для проверки гипотезы Γ_1 против альтернативы Γ_2

T_1. Статистика отношения максимумов правдоподобий (ОтМП)

Пусть $y = (y_1, \ldots, y_n)$, y_i – н.о.р. случайные величины, $i = \overline{1,n}$,

$$\mathcal{L}'(y) = p(u; \theta_0), \tag{15}$$

$u \in Y$, $\theta_0 = (\theta_{01}, \theta_{02}, \ldots, \theta_{0m})^T \in \Theta \subset \mathsf{R}^m$, гипотеза Γ_1 : $\theta_{0r} = a_r$, где a_r – известные величины, $r = \overline{1,k}$, а значение параметров θ_{0r} для $r \geq k+1$ – неизвестны. Здесь k – целое положительное число, $k < m$.

Далее при гипотезе Γ_2 : $\theta_{0r} = a_r$ для $r = \overline{1,k}$, и плотности распределения $p(u;\theta_0)$, удовлетворяющей условиям регулярности теоремы 1 раздела "Проверка сложных гипотез на основе статистики отношения максимумов правдоподобия", статистика критерия

$$T_{\text{ОтМП}}(y) = 2 \ln \frac{\max\limits_{\theta \in \Theta} p(y;\theta)}{\max\limits_{\theta \in \Theta : \theta_{0r} = a_r, r = \overline{1,k}} p(y;\theta)}. \tag{16}$$

в задаче проверки гипотезы Γ_1 против альтернативы Γ_2 при $n \to \infty$

$$T_{\text{ОтМП}}(y) = \chi_k^2 + o_d(1). \tag{17}$$

■

T_2. Статистика хи-квадрат Пирсона

Пусть случайный вектор

$$\nu = (\nu_1, \ldots, \nu_r)^T, \quad \nu_j \geq 0 \qquad \text{для } j = \overline{1,r}, \quad \sum_{j=1}^{r} \nu_j = n, \tag{18}$$

имеет полиномиальное распределение, т.е.

$$\mathcal{L}'(\nu) = \prod_{j=1}^{r} \frac{(p_j(\theta_0))^{\nu_j}}{\nu_j!}, \tag{19}$$

где набор вероятностей $p_1(\theta_0), \ldots, p_r(\theta_0)$ таков, что $p_j(\theta_0) > 0$ для всех $j = \overline{1,r}$, $\sum\limits_{j=1}^{r} p_j(\theta_0) = 1$, $\theta_0 \in \Theta$, $\Theta \subset \mathsf{R}^m$.

Статистика критерия хи-квадрат Пирсона имеет вид

$$X^2(\tilde{\theta}_n) = \sum_{j=1}^{r} \frac{\left(\nu_j - np_j(\tilde{\theta}_n)\right)^2}{np_j(\tilde{\theta}_n)}, \tag{20}$$

где

$$\tilde{\theta}_n = Arg \min_{\theta \in \Theta} X^2(\theta), \tag{21}$$

или, что эквивалентно при $n \to \infty$,

$$\tilde{\theta}_n = Arg \max_{\theta \in \Theta} \prod_{j=1}^{r} \frac{(p_i(\theta))^{\nu_i}}{\nu_j!}. \tag{22}$$

Известно, что при $n \to \infty$

$$X^2(\tilde{\theta}_n) \stackrel{d}{=} X^2_{r-1-m} + o_d(1). \tag{23}$$

■

T_3. Статистика хи-квадрат в форме Никулина-Джапаридзе

Пусть случайный вектор

$$\nu = (\nu_1, \dots, \nu_r)^T, \quad \nu_j \geq 0 \quad \text{для } j = \overline{1,r}, \quad \sum_{j=1}^{r} \nu_j = n, \tag{24}$$

имеет полиномиальное распределение, т.е.

$$\mathcal{L}'(\nu) = \prod_{j=1}^{r} \frac{(p_j(\theta_0))^{\nu_j}}{\nu_j!}, \tag{25}$$

где вектор вероятностей $\mathbf{p}(\theta_0) = (p_1(\theta_0), \dots, p_r(\theta_0))^T$ таков, что $p_j(\theta_0) > 0$, $j = \overline{1,r}$, $\sum\limits_{j=1}^{r} p_j(\theta_0) = 1$, $\theta_0 \in \Theta$, $\Theta \subset \mathsf{R}^m$.

Для $\hat{\theta}_n$ – оценки параметра θ_0 статистика критерия хи-квадрат Пирсона имеет вид

$$X^2(\hat{\theta}_n) = \sum_{j=1}^{r} \frac{\left(\nu_j - np_j(\hat{\theta}_n)\right)^2}{np_j(\hat{\theta}_n)}. \tag{26}$$

Далее предположим, что для всех $\theta \in \Theta$, $\theta = (\theta_1, \dots, \theta_m)^T$, выполнены также следующие условия

$$\mathcal{Y}_1 \quad : \quad \frac{\partial^2 p_j(\theta)}{\partial\theta_k \partial\theta_s} - \text{ непрерывные функции для } k, s = \overline{1,m};$$

$$\mathcal{Y}_2 \quad : \quad \text{матрица } A(\theta) = \left(\frac{\partial p_j(\theta)}{\partial \theta_k}; j = \overline{1,r}, k = \overline{1,m}\right) \text{ имеет ранг } \quad m.$$

Введем также следующие величины:

вектор

$$L(\theta) = (l_1(\theta), \dots, l_m(\theta))^m, \quad \text{где} \quad l_j(\theta) = \sum_{i=1}^{r} \frac{\nu_i}{np_i(\theta)} \frac{\partial p_i(\theta)}{\partial \theta_j}, \quad j = \overline{1,m}, \tag{27}$$

и матрицу

$$J(\theta) = (J_{ks}(\theta), \quad k, s = \overline{1,m}), \quad \text{где} \tag{28}$$

$$J_{ks}(\theta) = \sum_{i=1}^{r} \frac{1}{p_i(\theta)} \frac{\partial p_i(\theta)}{\partial \theta_k} \frac{\partial p_i(\theta)}{\partial \theta_s}, \quad k, s = \overline{1,m}. \tag{29}$$

Имеет место следующее утверждение: если оценка $\hat{\theta}_n$ является $\sqrt{n}$-состоятельной, то статистика

$$W^2(\tilde{\theta}_n) = X^2(\tilde{\theta}_n) - nL^T(\tilde{\theta}_n)J^{-1}(\tilde{\theta}_n)L(\tilde{\theta}_n) \stackrel{d}{=} \chi^2_{r-1-m} + o_d(1), \tag{30}$$

см. [25]. ■

T_4. Статистика, основанная на точечной оценке параметра распределения Хоффмана

Пусть $y = (y_1, \ldots, y_n)^T$, y_i – н.о.р. случайные величины, $i = \overline{1, n}$,

$$y_1 \stackrel{d}{=} HOFM(\alpha_0, \beta_0, \gamma_0), \tag{31}$$

$\theta_0 = (\alpha_0, \beta_0, \gamma_0)$, $\hat{\theta}_n$ – состоятельная при $n \to \infty$ оценка параметра θ_0.

Пусть гипотеза $\Gamma_1 : \gamma_0 = \gamma_1$, где γ_1 – известная величина, $\Gamma_2 : \gamma_0 \neq \gamma_1$. Далее, $\hat{\gamma}_n$ – состоятельная $\sqrt{n}$-асимптотически нормальная асимптотически несмещенная оценка параметра γ_0, для которой $B(\theta_0)$ является дисперсией предельного при $n \to \infty$ распределения. Тогда

$$T_4(y) = \sqrt{n}\left|\frac{\hat{\gamma}_n - \gamma_0}{\sqrt{B(\theta_0)}}\right| \stackrel{d}{=} |N(0,1)| + o_d(1). \tag{32}$$

■