Сведения из ТВ и теории оценивания

Олег Николаевич Граничин

Санкт-Петербургский государственный университет,
математико-механический факультет

20 февраля 2013

Стохастичность

Греческое слово

- целиться, метить (целиться во что (в кого)-л.)
- стремиться к тому, чтобы судьями были самые влиятельные люди
- применяться, приспособляться
- умозаключать, судить, догадываться, разгадывать (догадываться о том, что требуется; заключать на основании чего-л.; путем догадок)

Случайные величины I

Пусть $(\Omega, \mathscr{F})$ — некоторое измеримое пространство и $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ — числовая прямая с системой борелевских множеств $\mathscr{B}(\mathbb{R})$. Действительная функция $\xi = \xi(\omega)$, определенная на $(\Omega, \mathscr{F})$ называется *$\mathscr{F}$-измеримой функцией или случайной величиной*, если для любого $B \in \mathscr{B}(\mathbb{R})$

$$\{\omega : \xi(\omega) \in B\} \in \mathscr{F}.$$

Пусть $(\Omega, \mathscr{F}, \mathbf{P})$ — произвольное вероятностное пространство. *Математическим ожиданием* $\mathrm{E}\{\xi\}$ произвольной случайной величины ξ называется интеграл Лебега от $\mathscr{F}$-измеримой функции $\xi = \xi(\omega)$ по мере $\mathbf{P}$, для которого (наряду с $\mathrm{E}\{\xi\}$) используются также следующие обозначения: $\int_\Omega \xi(\omega)\mathbf{P}\{d\omega\}$ или $\int_\Omega \xi\, d\mathbf{P}$. *Дисперсией* случайной величины ξ называется величина $\sigma^2 = \mathrm{E}\{(\xi - \mathrm{E}\{\xi\})^2\}$, при этом величина $\sigma > 0$ называется *стандартным отклонением*.

Случайные величины II

Ковариация пары случайных величин ζ и η:

$$\mathrm{cov}\{\zeta,\eta\} = \mathrm{E}\{(\zeta - \mathrm{E}\{\zeta\})(\eta - \mathrm{E}\{\eta\})\}.$$

Если $\mathrm{cov}\{\zeta,\eta\} = 0$, то говорят, что случайные величины ζ и η *не коррелированы*.

Распределением случайной величины ξ на $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ называется вероятностная мера $\mathrm{P}_\xi(\cdot)$ на $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$:

$$\mathrm{P}_\xi(B) = \mathbf{P}\{\omega : \xi(\omega) \in B\},\ B \in \mathscr{B}(\mathbb{R}).$$

Из определения следует, что $\mathrm{E}\{\xi\} = \int_{\mathbb{R}} x \mathrm{P}_\xi(dx)$.

Функция распределения случайной величины ξ

$$P_\xi(x) = \mathbf{P}\{\omega : \xi(\omega) \le x\},\ x \in \mathbb{R}.$$

Неотрицательная функция $p_\xi(\cdot)$ называется *плотностью* функции распределения случайной величины ξ, если

$$P_\xi(x) = \int_{-\infty}^{x} p_\xi(t)dt.$$

Случайные величины III

Случайная величина ξ называется *гауссовской* (или *нормально распределенной*) с параметрами M и σ^2 ($\xi \sim \mathcal{N}(M,\sigma^2)$), $|M| < \infty$, $\sigma > 0$, если её плотность $p_\xi(\cdot)$ имеет следующий вид:

$$p_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-M)^2}{2\sigma^2}}.$$

Случайные величины $\xi_1,\ldots,\xi_n$ называются *независимыми* (*независимыми в совокупности*), если для любых $B_1,\ldots,B_n \in \mathscr{B}(\mathbb{R})$

$$\mathbf{P}\{\xi_1 \in B_1,\ldots,\xi_n \in B_n\} = \mathbf{P}\{\xi_1 \in B_1\}\cdots\mathbf{P}\{\xi_n \in B_n\}.$$

Пусть ξ и η — независимые случайные величины с $\mathrm{E}\{|\xi|\} < \infty$ и $\mathrm{E}\{|\eta|\} < \infty$. **Тогда** $\mathrm{E}\{|\xi\eta|\} < \infty$ и

$$\mathrm{E}\{\xi\eta\} = \mathrm{E}\{\xi\}\mathrm{E}\{\eta\}.$$

Понятие случайная величина естественным образом обобщается и на векторный случай.

Некоторые неравенства для случайных величин

Неравенство Маркова *Пусть* ξ *— неотрицательная случайная величина.* **Тогда** *для всякого* $\varepsilon > 0$

$$\mathsf{P}\{\xi \geq \varepsilon\} \leq \frac{\mathrm{E}\{\xi\}}{\varepsilon}.$$

Неравенство Чебышёва *Пусть* ξ *— случайная величина.* **Тогда** *для всякого* $\varepsilon > 0$

$$\mathsf{P}\{|\xi| \geq \varepsilon\} \leq \frac{\mathrm{E}\{|\xi|^2\}}{\varepsilon^2}.$$

Неравенство Иенсена *Пусть* $g(x)$ *— выпуклая функция, а* ξ *— случайная величина с* $\mathrm{E}\{\xi\} < \infty$. **Тогда**

$$g(\mathrm{E}\{\xi\}) \leq \mathrm{E}\{g(\xi)\}.$$

Неравенство Гёльдера *Пусть* $1 < p < \infty$, $1 < q < \infty$ *и* $\frac{1}{p} + \frac{1}{q} = 1$. *Если* $\mathrm{E}\{|\xi|^p\} < \infty$ *и* $\mathrm{E}\{|\eta|^q\} < \infty$, **то** $\mathrm{E}\{|\xi\eta|\} < \infty$ *и*

$$\mathrm{E}\{|\xi\eta|\} < (\mathrm{E}\{|\xi|^p\})^{1/p}(\mathrm{E}\{|\eta|^q\})^{1/q}.$$

Закон больших чисел для независимых случайных величин

Закон больших чисел *Пусть* $\xi_1, \xi_2, \ldots$ *— последовательность независимых одинаково распределенных случайных величин с* $\mathrm{E}\{\xi_1\} < \infty$, $S_n = \xi_1 + \ldots + \xi_n$ *и* $\mathrm{E}\{\xi_1\} = M$.
Тогда *при* $n \to \infty$

$$\forall \varepsilon > 0 \ \mathbf{P}\{|\frac{S_n}{n} - M| \geq \varepsilon\} \to 0.$$

Теорема Кантелли *Пусть* $\xi_1, \xi_2, \ldots$ *— последовательность независимых случайных величин с конечным четвертым моментом:*

$$\mathrm{E}\{|\xi_n - \mathrm{E}\{\xi_n\}|^4\} \leq C < \infty, \ n \geq 1,$$

$S_n = \xi_1 + \ldots + \xi_n$.
Тогда *при* $n \to \infty$ *с вероятностью единица*

$$\frac{S_n - \mathrm{E}\{S_n\}}{n} \to 0.$$

Усиленный закон больших чисел (Колмогорова) *Пусть* $\xi_1, \xi_2, \ldots$ *— последовательность независимых случайных величин с конечными вторыми моментами, положительные числа* β_n *таковы, что*

$$\beta_n \to \infty, \sum \frac{\mathrm{E}\{(\xi_n - \mathrm{E}\{\xi_n\})^2\}}{\beta_n^2} < \infty,$$

$S_n = \xi_1 + \ldots + \xi_n$.
Тогда *при* $n \to \infty$ *с вероятностью единица*

$$\frac{S_n - \mathrm{E}\{S_n\}}{\beta_n} \to 0.$$

Неравенство Хёффдинга

Hoeffding's inequality. *Пусть* $\xi_1, \xi_2, \ldots$ — *последовательность независимых случайных ограниченных величин, принимающих значения в интервалах* $[\alpha_k, \beta_k]$, *и* $S_n = \xi_1 + \ldots + \xi_n$. **Тогда** *для всякого* $\varepsilon > 0$

$$\mathbf{P}\left\{\frac{S_n - \mathrm{E}\{S_n\}}{n} \le -\varepsilon\right\} \le e^{-\frac{2n^2\varepsilon^2}{\sum_{k=1}^{n}(\beta_k-\alpha_k)^2}}$$

и

$$\mathbf{P}\left\{\frac{S_n - \mathrm{E}\{S_n\}}{n} \ge \varepsilon\right\} \le e^{-\frac{2n^2\varepsilon^2}{\sum_{k=1}^{n}(\beta_k-\alpha_k)^2}}.$$

Следствие. Если $[\alpha_k, \beta_k] = [0, 1]$, **то**

$$\mathbf{P}\left\{\frac{|S_n - \mathrm{E}\{S_n\}|}{n} \le \varepsilon\right\} \le 2e^{-2n\varepsilon^2}.$$

Наилучшая аппроксимация одной случайной величины с помощью другой

Пусть ζ и η — произвольные случайные величины (векторы), принимающие значения соответственно в $\mathbb{R}$ и $\mathbb{R}^s$, определенные на некотором вероятностном пространстве Ω, и $\mathscr{G}$ — некоторое семейство функций, отображающих $\mathbb{R}^s$ в $\mathbb{R}$, заданных с точностью до конечномерного набора параметров, называемое *регрессионной моделью*. *Требуется* найти функцию $g(\cdot) \in \mathscr{G}$, минимизирующую

$$\mathrm{E}\{\|\zeta - g(\eta)\|^2\}.$$

Если $\mathscr{G}$ — класс всех измеримых функций из $\mathbb{R}^s$ в $\mathbb{R}$, то соответствующей минимизирующей функцией $g(\cdot)$ является $g(\eta) = \mathrm{E}\{\zeta|\eta\}$ — условное (при условии η) среднее случайной величины ζ, называемое *регрессией* ζ *по* η.

Наиболее распространенной является *линейная регрессионная модель*, когда требуется найти наилучшую в среднеквадратичном смысле аппроксимацию случайной величины ζ с помощью линейной функции от случайной величины η.

Если вектор, составленный из компонент случайных величин ζ и η — гауссовский, то регрессия $\mathrm{E}\{\zeta|\eta\}$ случайной величины ζ по η совпадает с линейной регрессией.

Оценивание по конечному числу наблюдений

Если имеется выборочная последовательность $\varphi_1, \varphi_2, \ldots, \varphi_N$ реализаций случайной величины η, то в рамках линейной регрессионной модели реализации $y_1, y_2, \ldots, y_N$ случайной величины ζ удобно представить в виде

$$y_n = \varphi_n^{\mathrm{T}} \theta + v_n, \; n = 1, 2, \ldots, N,$$

где θ — вектор коэффициентов. В этом представлении невязки v_n, $n = 1, 2, \ldots, N$, интерпретируются как ошибки наблюдения. Обозначим $Y = Y^N = (y_1, y_2, \ldots, y_N)^{\mathrm{T}}$ — наблюдаемый в момент времени N вектор, являющийся функцией входных воздействий, помех в канале измерения и некоторого векторного параметра θ. *Требуется* по значению вектора Y получить хорошую оценку $\hat{\theta} = \hat{\theta}^N$ вектора θ.

Свойства оценок

- Оценка $\hat{\theta}$ называется *линейной*, если она имеет вид $\hat{\theta} = \Gamma Y$ с некоторой матрицей коэффициентов Γ.
- Оценка $\hat{\theta}$ называется *несмещенной*, если $\mathrm{E}\{\hat{\theta}\} = \theta$.
- Последовательность оценок $\{\hat{\theta}^N\}_{N=1}^{\infty}$ называется *состоятельной*, если для любого $\varepsilon > 0$

$$\lim_{N\to\infty} \mathsf{P}\{\|\hat{\theta}^N - \theta\|^2 > \varepsilon\} = 0,$$

- и называется *сильносостоятельной*, если с вероятностью единица

$$\lim_{N\to\infty} \hat{\theta}^N = \theta.$$

Мартингалы, супер- и субмартингалы

Последовательность случайных величин $v_0, \ldots, v_n, \ldots$: $\mathrm{E}\{|v_n|\} < \infty$, называется *мартингалом*, если

$$\mathrm{E}\{v_{n+1}|v_0, \ldots, v_n\} = v_n,$$

супермартингалом, если

$$\mathrm{E}\{v_{n+1}|v_0, \ldots, v_n\} \leq v_n,$$

субмартингалом, если

$$\mathrm{E}\{v_{n+1}|v_0, \ldots, v_n\} \geq v_n,$$

Теорема Дуба о сходимости супермартингалов

Theorem

Если последовательность $v_0, \ldots, v_n, \ldots$ *— супермартингал,* **тогда** *с вероятностью единица существует предел:* $v_\infty = \lim_{n\to\infty} v_n$, *и* $\mathrm{E}|v_\infty| < \infty$.

Последовательности случайных величин, близкие к супермартингалам. Лемма Роббинса-Зигмунда

Lemma

Если $v_0, \ldots, v_n, \ldots$ *— последовательность неотрицательных случайных величин:* $v_n \geq 0$, $\mathrm{E}\{v_0\} < \infty$ *и*

$$\mathrm{E}\{v_{n+1} | v_0, \ldots, v_n\} \leq (1 - \alpha_n) v_n + \beta_n,$$

$$0 \leq \alpha_n \leq 1,\ \beta_n \geq 0,\ \sum \alpha_n = \infty,\ \sum \beta_n < \infty,\ \frac{\beta_n}{\alpha_n} \to 0,$$

тогда *с вероятностью единица* $v_n \to 0$, $\mathrm{E}\{v_n\} \to 0$ *и*

$$\forall \varepsilon > 0,\ n > 0\ \ \mathbf{P}\{v_j \leq \varepsilon\ \forall j \geq n\} \geq 1 - \varepsilon^{-1}(\mathrm{E}\{v_n\} + \sum_{i=1}^{\infty} \beta_i).$$

Лемма Чжуна

Lemma

Если последовательность неотрицательных чисел $\{u_n\}$, $u_n \geq 0$, *удовлетворяет неравенству:*

$$u_{n+1} \leq (1 - \frac{c}{n})u_n + \frac{d}{n^{p+1}}, \ \ d > 0, \ p > 0, \ c > p,$$

тогда $u_n \leq \frac{d}{c-p} n^{-p} + o(n^{-p})$ *при* $c > p$,
$u_n = \mathcal{O}(n^{-c} \ln n)$ *при* $c = p$,
$u_n = \mathcal{O}(n^{-c})$ *при* $c < p$.

Рассмотрим дискретную стохастическую систему

$$\bar{x}_{t+1} = \bar{x}_t + F_t(\alpha_t, \bar{x}_t, \bar{w}_t), \qquad t = 0, 1, 2, \ldots, \tag{1}$$

где $\bar{x}_t \in \mathbb{R}^n$ — вектор состояний, $\bar{w}_t \in \mathbb{R}^m$ — случайный вектор возмущений, α_t — параметр шага, $F_t(\cdot,\cdot,\cdot) : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ — некоторые вектор-функции от параметра шага α_t и двух аргументов.
Обычно вектор-функции $F_t(\alpha_t, \bar{x}_t, \bar{w}_t)$ имеют вид:
$F_t(\alpha_t, \bar{x}_t, \bar{w}_t) = \alpha_t \bar{F}_t(\bar{x}_t, \bar{w}_t)$.
Будем считать, что при $k \to \infty$ $\alpha_t \to \alpha \geq 0$.

Усредненные модели

Выпишем соответствующую (1) усредненную дискретную модель

$$\bar{z}_{t+1} = \bar{z}_t + G_t(\alpha_t, \bar{z}_t),\ \bar{z}_0 = \bar{x}_0, \tag{2}$$

в которой функции $G_t(\cdot,\cdot) : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ при детерминированных значениях $\bar{z}$ определены следующим образом:

$$G_t(\alpha_t, \bar{z}) = \mathrm{E} F_t(\alpha_t, \bar{z}, \bar{w}_t). \tag{3}$$

Предположив, что для любого $\bar{z} \in \mathbb{R}^n$ существует предел

$$R(\alpha, \bar{z}) = \lim_{t\to\infty} \frac{1}{\alpha_t} G_t(\bar{z}) = \lim_{t\to\infty} \frac{1}{\alpha_t} \mathrm{E} F_t(\bar{z}, \bar{w}_t), \tag{4}$$

наряду с (1) рассмотрим соответствующую непрерывную систему (непрерывную модель):

$$\frac{d\bar{x}}{d\tau} = R(\alpha, \bar{x}), \tag{5}$$

в которой траектории — это вектор-функции $\bar{x}(\tau) \in \mathbb{R}^n$, зависящие от "непрерывного" аргумента $\tau \in \mathbb{R}$, $\tau \geq 0$.

"Близость" траекторий

- Если параметры шага α_t достаточно малы ($\alpha_t \le \alpha$), тогда траектории $\{\bar{x}_t\}$ системы (1) близки к точкам $\{\bar{x}(\tau_t)\}$ траекторий системы (5) при $\tau_t = \alpha_0 + \dots + \alpha_{t-1}$.
- Если параметры шага α_t стремятся к нулю при $t \to \infty$, то некоторые асимптотические свойства решений системы (1) (например, устойчивость, предельная ограниченность и т. п.) могут быть аналогичны свойствам решений непрерывной модели (5).

В случае близости систем (1) и (5) в указанном выше смысле для целей системного анализа и проектирования можно использовать упрощенную модель (5) вместо (1). Такой подход получил название *метод непрерывных моделей* (в англоязычной литературе *ODE approach* или *Derevitskii-Fradkov-Ljung (DFL) - scheme*.