Лекция 11

Функциональная предельная теорема для эмпирического процесса

Пусть $\eta_1, \eta_2, \ldots$ – независимые одинаково распределенные случайные величины с функцией распределения $F(t)$. Введем для каждого $n \in \mathbf{N}$ *эмпирическую функцию распределения*

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} I_{\{\eta_i \le t\}}, \qquad t \in \mathbb{R},$$

где $I_A(\cdot)$ означает индикатор случайного события A. Ясно, что $F_n(t)$ – случайная величина, и по усиленному закону больших чисел почти наверное

$$\lim_{n\to\infty} F_n(t) = F(t).$$

Введем так называемый *эмпирический процесс*:

$$Z_n(t) = \sqrt{n}\left(F_n(t) - F(t)\right), \qquad t \in \mathbb{R}.$$

Иногда рассматривается частный случай этого процесса, когда случайные величины $\eta_1, \eta_2, \ldots$ имеют равномерное распределение на интервале $(0,1)$. В этом случае $F(t) = t$ при $t \in [0,1]$ и, следовательно, $Z_n(t) = \sqrt{n}\left(F_n(t) - t\right)$ при $t \in [0,1]$. Следующий результат принадлежит Донскеру.

Теорема 1. *Если $\eta_1, \eta_2, \ldots$ – независимые случайные величины, имеющие равномерное распределение на интервале $(0,1)$, то при $n \to \infty$*

$$\{Z_n(t),\ t \in [0,1]\} \xrightarrow{D} W_0,$$

где W_0 – броуновский мост, знак $\xrightarrow{D}$ означает сходимость по распределению в пространстве $D[0,1]$.

Сначала установим полезную теорему, имеющую применения и вне данного контекста. Пусть $\eta_1, \eta_2, \ldots$ – независимые случайные величины, имеющие равномерное распределение на интервале $(0,1)$. Положим при $t \in [0,1]$

$$Y_n(t) = \sum_{i=1}^{n} I_{\{\eta_i \le t\}}, \qquad \widetilde{Y}_n(t) = Y_n\left(\frac{\lfloor nt \rfloor}{n}\right).$$

Рассмотрим последовательность независимых случайных величин X_i, $i \in \mathbb{N}$, имеющих распределение Пуассона с параметром $a > 0$, и положим $S_n = X_1 + \ldots + X_n$, $n \in \mathbb{N}$.

Теорема 2. *Конечномерные распределения процессов*

$$\left\{\widetilde{Y}_n(t),\ t \in [0,1]\right\} \text{ и } \left\{S_{\lfloor nt \rfloor},\ t \in [0,1] \ \middle|\ S_n = n\right\}$$

совпадают.

Доказательство теоремы 2. Предположим, что $m \in \mathbb{N}$ и $p_1, \ldots, p_m$ – неотрицательные числа, такие что $p_1 + \ldots + p_m = 1$. Напомним, что случайный вектор $\{\xi_1, \ldots, \xi_m\}$, $m \in \mathbb{N}$, имеет полиномиальное распределение с параметрами n и $p_1, \ldots, p_m$, если для произвольных целых неотрицательных чисел $k_1, \ldots, k_m$ таких, что $\sum_{i=1}^m k_i = n$,

$$\mathbf{P}\left(\xi_1 = k_1, \ldots, \xi_m = k_m\right) = \frac{n!}{k_1! \ldots k_m!} p_1^{k_1} \ldots p_m^{k_m}.$$

Лемма 1. *Пусть $m \in \mathbb{N}$ и $0 = t_0 < t_1 < \ldots < t_m < t_{m+1} = 1$. Тогда случайный вектор $\{Y_n(t_i) - Y_n(t_{i-1}),\ i \in\ I_{m+1}\}$ имеет полиномиальное распределение с параметрами n и $p_1, \ldots, p_{m+1}$,* где $p_i = t_i - t_{i-1}$, $i \in\ I_{m+1}$.

Доказательство. Рассмотрим случайную величину η, имеющую равномерное распределение на интервале $(0,1)$. Будем говорить, что достигается успех i-го типа, если $t_{i-1} < \eta \le t_i$, $i \in\ I_{m+1}$. Тогда $Y_n(t_i) - Y_n(t_{i-1})$ – число успехов i-го типа в n независимых испытаниях, $i \in\ I_{m+1}$. Хорошо известно, что случайный вектор, состоящий из таких компонент имеет полиномиальное распределение. Лемма доказана.

Следствие 1. *Пусть $m \in \mathbb{N}$ и $0 = t_0 < t_1 < \ldots < t_m < t_{m+1} = 1$. Тогда случайный вектор $\left\{\widetilde{Y}_n(t_i) - \widetilde{Y}_n(t_{i-1}),\ i \in\ I_{m+1}\right\}$ имеет полиномиальное распределение с параметрами n и $\widetilde{p}_1, \ldots, \widetilde{p}_{m+1}$, где $\widetilde{p}_i = \left(\lfloor nt_i \rfloor - \lfloor nt_{i-1} \rfloor\right)/n$, $i \in\ I_{m+1}$.*

Лемма 2. *Пусть $m, n \in \mathbb{N}$ и натуральные числа $n_1, n_2, \ldots, n_m$ возрастают. Тогда случайный вектор $\left\{ S_{n_1}, S_{n_2} - S_{n_1}, \ldots, S_{n_m} - S_{n_{m-1}} \;\middle|\; S_{n_m} = n\right\}$ имеет полиномиальное распределение с параметрами n и $p_1, \ldots, p_m$, где $p_i = \left(n_i - n_{i-1}\right)/n_m$, $i \in\ I_m$, причем $n_0 = 0$.*

Доказательство. При целых неотрицательных числах $k_1, \ldots, k_m$ таких, что $\sum_{i=1}^m k_i = n$, находим, что

$$\mathbf{P}\left(S_{n_i} - S_{n_{i-1}} = k_i,\ i \in I_m \;\middle|\; S_{n_m} = n\right) =$$

$$= \frac{1}{\mathbf{P}\left(S_{n_m} = n\right)} \prod_{i=1}^m \mathbf{P}\left(S_{n_i} - S_{n_{i-1}} = k_i\right) =$$

$$= \frac{n!}{\left(an_m\right)^n e^{-an_m}} \prod_{i=1}^m \frac{\left(a\left(n_i - n_{i-1}\right)\right)^{k_i} e^{-a(n_i - n_{i-1})}}{k_i!} =$$

$$= \frac{n!}{k_1! \ldots k_m!} \prod_{i=1}^m \left(\frac{n_i - n_{i-1}}{n_m}\right)^{k_i} = \frac{n!}{k_1! \ldots k_{m+1}!} \prod_{i=1}^m p_i^{k_i}.$$

Лемма доказана.

Следствие 2. *Пусть $m \in \mathbb{N}$ и $0 = t_0 < t_1 < \ldots < t_m < t_{m+1} = 1$. Тогда случайный вектор $\left\{ S_{\lfloor nt_1 \rfloor}, S_{\lfloor nt_2 \rfloor} - S_{\lfloor nt_1 \rfloor}, \ldots, S_{\lfloor nt_{m+1} \rfloor} - S_{\lfloor nt_m \rfloor} \;\middle|\; S_n = n\right\}$ имеет полиномиальное распределение с параметрами n и $\widetilde{p}_1, \ldots, \widetilde{p}_{m+1}$, где $\widetilde{p}_i = \left(\lfloor nt_i \rfloor - \lfloor nt_{i-1} \rfloor\right)/n$, $i \in\ I_{m+1}$.*

Из следствий 1 и 2 следует утверждение теоремы 2.

Доказательство теоремы 1. Напомним, что при $t \in [0,1]$

$$Z_n(t) = \sqrt{n}\left(F_n(t) - t\right).$$

Поскольку $F_n(t) = Y_n(t)/n$, то

$$Z_n(t) = \frac{Y_n(t) - nt}{\sqrt{n}}. \tag{1}$$

Положим

$$\widetilde{Z}_n(t) = Z_n\left(\frac{\lfloor nt \rfloor}{n}\right).$$

Из (1) вытекает, что

$$\widetilde{Z}_n(t) = \frac{\widetilde{Y}_n(t) - \lfloor nt \rfloor}{\sqrt{n}}. \tag{2}$$

Наша задача показать, что при $n \to \infty$

$$Z_n \xrightarrow{D} W_0. \tag{3}$$

Докажем сначала, что при $n \to \infty$

$$\widetilde{Z}_n \xrightarrow{D} W_0. \tag{4}$$

Ввиду теоремы 2 совпадают конечномерные распределения процессов

$$\left\{\widetilde{Y}_n(t),\ t \in [0,1]\right\} \text{ и } \left\{ S_{\lfloor nt \rfloor},\ t \in [0,1] \;\middle|\; S_n = n \right\}.$$

В дальнейшем будем рассматривать частный случай второго из этих процессов, когда $a = 1$. Положим $X_i^* = X_i - 1$ при $i \in \mathbb{N}$ и

$$S_n^* = X_1^* + \ldots + X_n^*, \quad n \in \mathbb{N}.$$

Поскольку $S_n^* = S_n - n$ при $n \in \mathbb{N}$, то

$$\left\{ S_{\lfloor nt \rfloor},\ t \in [0,1] \;\middle|\; S_n = n \right\} = \left\{ S^*_{\lfloor nt \rfloor} + \lfloor nt \rfloor,\ t \in [0,1] \;\middle|\; S_n^* = 0 \right\}.$$

Таким образом, совпадают конечномерные распределения процессов

$$\left\{\widetilde{Y}_n(t),\ t \in [0,1]\right\} \text{ и } \left\{ S^*_{\lfloor nt \rfloor} + \lfloor nt \rfloor,\ t \in [0,1] \;\middle|\; S_n^* = 0 \right\}$$

и, значит (см. (2)), совпадают конечномерные распределения процессов

$$\left\{\widetilde{Z}_n(t),\ t \in [0,1]\right\} \text{ и } \left\{ S^*_{\lfloor nt \rfloor}/\sqrt{n},\ t \in [0,1] \;\middle|\; S_n^* = 0 \right\}.$$

Заметим, что $\mathbf{E}X_1^* = 0$, $\mathbf{E}\left(X_1^*\right)^2 = 1$ и, следовательно, по функциональной предельной теореме Лиггетта при $n \to \infty$

$$\left\{ S^*_{\lfloor nt \rfloor}/\sqrt{n},\ t \in [0,1] \;\middle|\; S_n^* = 0 \right\} \xrightarrow{D} W_0.$$

Таким образом, справедливо соотношение (4).

Лемма 3. *При* $n \to \infty$

$$\rho_{\text{ск}} \left(Z_n, \widetilde{Z}_n \right) \xrightarrow{P} 0.$$

Доказательство. При фиксированном $\delta > 0$ рассмотрим функционал $x \to w_x(\delta)$, где $x \in D[0,1]$. Он равномерно непрерывен относительно равномерной метрики при всех $x \in D[0,1]$ и, следовательно (см. лемму 1 лекции 8.1), он непрерывен относительно метрики Скорохода при всех $x \in C[0,1]$. Поскольку траектории броуновского моста почти наверное непрерывны, то при каждом $\delta > 0$ (см. теорему $1'$ лекции 4) при $n \to \infty$

$$w_{\widetilde{Z}_n}(\delta) \xrightarrow{D} w_{W_0}(\delta)$$

(по распределению в $D[0,1]$). Поэтому для всех $\varepsilon > 0$ (за исключением не более чем счетного множества)

$$\limsup_{n\to\infty} \mathbf{P} \left(w_{\widetilde{Z}_n}(\delta) \geq \varepsilon \right) = \mathbf{P} \left(w_{W_0}(\delta) \geq \varepsilon \right)$$

и, значит, для каждого $\varepsilon > 0$

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \mathbf{P} \left(w_{\widetilde{Z}_n}(\delta) \geq \varepsilon \right) = \lim_{\delta\to 0} \mathbf{P} \left(w_{W_0}(\delta) \geq \varepsilon \right) = 0. \tag{5}$$

Из (5) следует, что для каждого $\varepsilon > 0$

$$\limsup_{n\to\infty} \mathbf{P} \left(w_{\widetilde{Z}_n}(3/n) \geq \varepsilon \right) = 0. \tag{6}$$

Из определения $\widetilde{Z}_n$ вытекает, что

$$w_{Z_n}(2/n) \leq w_{\widetilde{Z}_n}(3/n). \tag{7}$$

Из (6) и (7) находим, что при $n \to \infty$

$$\limsup_{n\to\infty} \mathbf{P} \left(w_{Z_n}(2/n) \geq \varepsilon \right) = 0. \tag{8}$$

Процесс $\widetilde{Z}_n$ является ступенчатым $(1/n)$-приближением процесса Z_n, поэтому (см. леммы 2 и 3 лекции 6)

$$\rho_{\text{ск}} \left(Z_n, \widetilde{Z}_n \right) \leq w_{Z_n}(2/n) + 1/n. \tag{9}$$

Из соотношений (8) и (9) следует, что

$$\limsup_{n\to\infty} \mathbf{P} \left(\rho_{\text{ск}} \left(Z_n, \widetilde{Z}_n \right) \geq \varepsilon \right) = 0.$$

Лемма доказана.

Из соотношения (4) и леммы 3 следует (3). Теорема 1 доказана.

В математической статистике известна следующая теорема Гливенко-Кантелли: почти наверное

$$\lim_{n\to\infty}\sup_{t\in\mathbb{R}}\left|F_n(t)-F(t)\right|=0.$$

Важно понять, с какой скоростью величина $\sup_{t\in\mathbb{R}}\left|F_n(t)-F(t)\right|$, получившая название *статистики Колмогорова*, стремится к нулю. Следующий результат, установленный А.Н. Колмогоровым, является одной из *фундаментальных теорем математической статистики.*

Теорема 3. *Пусть* $\eta_1,\eta_2,\ldots$ *– независимые одинаково распределенные случайные величины с непрерывной функцией распределения* $F(t)$, $t\in\mathbb{R}$. *Тогда для произвольного* $x\geq 0$

$$\lim_{n\to\infty}\mathbf{P}\left(\sqrt{n}\sup_{t\in\mathbb{R}}\left|F_n(t)-F(t)\right|\leq x\right)=K(x),$$

где $K(\cdot)$ *– функция распределения Колмогорова, т.е.*

$$K(0)=0,\qquad K(x)=1+2\sum_{k=1}^{\infty}(-1)^k\exp\left(-2k^2x^2\right),\quad x>0.$$

Доказательство. Сначала установим следующий полезный результат.

Лемма 4. *Если* $F(t)$ *– непрерывная функция распределения, то случайная величина* $\sup_{t\in\mathbb{R}}\left|F_n(t)-F(t)\right|$ *имеет такое же распределение, как* $\sup_{t\in[0,1]}\left|\widetilde{F}_n(t)-t\right|$, *где* $\widetilde{F}_n(t)$ *– эмпирическая функция распределения, построенная по* $\widetilde{\eta}_1,\widetilde{\eta}_2,\ldots$ *– независимым одинаково распределенным случайным величинам, имеющим равномерное распределение на интервале* $(0,1)$.

Доказательство. Для простоты изложения будем считать, что функция распределения $F(t)$, $t\in\mathbb{R}$, строго возрастает, тогда существует обратная функция $F^{-1}(y)$, $y\in(0,1)$. Случайные величины $\widetilde{\eta}_i=F(\eta_i)$, $i\in\mathbb{N}$, имеют равномерное распределение на интервале $(0,1)$. Действительно, при $y\in(0,1)$

$$\mathbf{P}\left(\widetilde{\eta}_i\leq y\right)=\mathbf{P}\left(F(\eta_i)\leq y\right)=\mathbf{P}\left(\eta_i\leq F^{-1}(y)\right)=F\left(F^{-1}(y)\right)=y.$$

Далее,

$$\sup_{t\in\mathbb{R}}\left|F_n(t)-F(t)\right|=\sup_{t\in\mathbb{R}}\left|\frac{1}{n}\sum_{i=1}^{n}I_{\{\eta_i\leq t\}}-F(t)\right|=$$

$$=\sup_{t\in\mathbb{R}}\left|\frac{1}{n}\sum_{i=1}^{n}I_{\{F(\eta_i)\leq F(t)\}}-F(t)\right|=\sup_{u\in[0,1]}\left|\frac{1}{n}\sum_{i=1}^{n}I_{\{\widetilde{\eta}_i\leq u\}}-u\right|.$$

Лемма доказана.

Продолжим доказательство теоремы 3. Покажем, что при $n\to\infty$

$$\sqrt{n}\sup_{t\in\mathbb{R}}\left|F_n(t)-F(t)\right|\xrightarrow{D}\sup_{t\in[0,1]}\left|W_0(t)\right|. \tag{10}$$

Рассмотрим отображение $f : x \to \sup_{t\in[0,1]} |x(t)|$, где $x \in D\,[0,1]$. Ранее было показано (см. доказательство теоремы 1 лекции 8.1), что это отображение равномерно непрерывно в равномерной метрике при всех $x \in D\,[0,1]$. Следовательно, по лемме 4 лекции 6 оно непрерывно относительно метрики Скорохода при всех $x \in C\,[0,1]$. Поскольку траектории броуновского моста почти наверное непрерывны, то из теоремы 1 следует (см. теорему $1'$ лекции 4), что если $\eta_1, \eta_2, \dots$ – независимые случайные величины, имеющие равномерное распределение на интервале $(0,1)$, то при $n \to \infty$

$$\sqrt{n} \sup_{t\in[0,1]} |F_n\,(t) - t| = \sup_{t\in[0,1]} |Z_n\,(t)| \stackrel{D}{\to} \sup_{t\in[0,1]} |W_0(t)|\,. \tag{11}$$

По лемме 3 левая часть соотношения (10) имеет такое же распределение, как и левая часть соотношения (11), поэтому из (11) следует (10).

Осталось вспомнить, что распределение правой части соотношения (10) является распределением Колмогорова (см теорему 3 лекции 9). Теорема 3 доказана.