§3. Оценка параметров. Метод моментов

Лекция 5

Мы приступаем к решению статистической проблемы оценки неизвестного значения параметра θ, индексирующего семейство $\mathcal{P} = \{P_\theta,\, \theta \in \Theta\}$ возможных распределений наблюдаемой случайной величины X. Будут рассматриваться только конечномерные параметрические пространства $\Theta = \mathrm{R}^k$, $k \geq 1$. Информация о значении θ поступает нам в виде выборочных данных $x^{(n)} = (x_1, \ldots, x_n)$ – результатов наблюдений n независимых копий $X^{(n)} = (X_1, \ldots, X_n)$ случайной величины X. Напомним, семейство $\mathcal{P}$ мы назвали вероятностной моделью, а случайный вектор $X^{(n)}$ – случайной выборкой объема n.

В этой проблеме, о которой мы несколько раз упоминали в предыдущем параграфе, пространство решений $\mathcal{D}$ совпадает с параметрическим пространством Θ, решающая функция $\delta = \delta(X^{(n)})$ – статистика с областью значений $\mathcal{T} = \Theta$ – называется *оценкой параметра* θ и обычно обозначается θ_n, $\hat{\theta}_n$, θ_n^* и тому подобное. Функции потерь $L(\theta,\, d)$ в проблеме оценивания обычно выбираются в виде неубывающей функции расстояния $|\, d - \theta\, |$ (в эвклидовой метрике) между значением оценки $d = \hat{\theta}_n(x^{(n)})$ и истинным значением θ оцениваемого параметра.

Основная задача статистической теории оценивания состоит в построении оценки $\theta_n^* = \theta_n^*(X^{(n)})$, *минимизирующей равномерно по* $\theta \in \Theta$ *функцию риска*

$$R(\theta;\, \hat{\theta}_n) = \mathbf{E}_\theta L(|\, \theta,\, \hat{\theta}_n(X^{(n)})\, |).$$

Таким образом, какова бы ни была статистическая оценка $\hat{\theta}_n$, для оценки θ_n^* с равномерно минимальным риском при любом $\theta \in \Theta$ справедливо неравенство $R(\theta;\, \theta_n^*) \leq R(\theta;\, \hat{\theta}_n)$.

Мы рассмотрим одно из решений этой задачи в случае оценки скалярного параметра ($\Theta = \mathrm{R}$) при квадратичной функции потерь $L(\theta, d) = (d - \theta)^2$, но сначала познакомимся с традиционно используемыми в статистической практике методами оценки параметров и изучим распределение этих оценок с целью вычисления их функции риска.

Конечно, далеко не все используемые на практике методы приводят к оптимальным оценкам, иногда бывает трудно найти оценку, обладающую хоть какими-нибудь привлекательными свойствами. Понятно, что считать оценкой любое измеримое отображение выборочного пространства $\mathcal{X}^n$ в параметрическое пространство Θ не совсем разумно, и поэтому мы введем некоторые условия, которым должна удовлетворять статистика $\hat{\theta}_n$, чтобы претендовать на роль *оценки*. Разрабатывая в дальнейшем методы оценивания и предлагая конкретные оценки, мы всегда будем проверять выполнимость этих условий.

Определение 3.1. Оценка $\hat{\theta}_n$ параметра θ называется *состоятельной*, если $\hat{\theta}_n(X^{(n)}) \underset{P}{\to} \theta$ при любом $\theta \in \Theta$, когда объем выборки $n \to \infty$. Оценка $\hat{\theta}_n$ называется *несмещенной в среднем*, если $\mathbf{E}_\theta\, \hat{\theta}_n(X^{(n)}) = \theta$, каково бы ни было значение $\theta \in \Theta$.

Напомним, что $\hat{\theta}_n(X^{(n)}) \underset{P}{\to} \theta$ означает, что для любого $\varepsilon > 0$

$$\lim_{n\to\infty} P_\theta \left(\left| \hat{\theta}_n(X^{(n)}) - \theta \right| > \varepsilon \right) = 0,$$

или, что то же

$$\lim_{n\to\infty} P_\theta \left(\left| \hat{\theta}_n(X^{(n)}) - \theta \right| \le \varepsilon \right) = 1. \tag{1}$$

Здесь, как обычно, в случае векторного параметра θ запись $|\,\theta_1 - \theta_2\,|$ означает расстояние между точками θ_1 и θ_2 эвклидова пространства Θ.

В предыдущем параграфе мы показали, что выборочные моменты $a_i = (1/n)\sum_1^n X_j^i$ являются состоятельными оценками соответствующих "теоретических" моментов $\alpha_i = \mathbf{E}_\theta X^i$, которые являются функциями оцениваемого параметра: $\alpha_i = \alpha_i(\theta),\ i = 1, 2, \ldots$. Этот результат указывает нам довольно простой метод построения состоятельных оценок в случае существования у распределения P_θ наблюдаемой случайной величины X момента порядка k, где k – число компонент $\theta_1, \ldots, \theta_k$ оцениваемого параметрического вектора θ.

Приравняем теоретические моменты выборочным и разрешим полученную таким образом систему уравнений $\alpha_i(\theta_1, \ldots, \theta_k) = a_i,\ i = 1, \ldots, k$ относительно переменных $\theta_1, \ldots, \theta_k$. Любое решение $\hat{\theta}_n(\mathbf{a}) =$

$(\hat{\theta}_{1n}(\mathbf{a}), \ldots, \hat{\theta}_{kn}(\mathbf{a}))$, $\mathbf{a} = (a_1, \ldots, a_n)$, этой системы называется *оценкой θ по методу моментов*, и прежде, чем исследовать свойства таких оценок, рассмотрим несколько примеров на применения *метода моментов*.

В курсе теории вероятностей, изучая новые вероятностные модели, мы всегда вычисляли их моментные характеристики. Например, мы знаем, что средние значения двухточечного распределения $\mathrm{B}(1,\,\theta)$, распределения Пуассона $\mathrm{P}(\theta)$ и показательного распределения $\mathrm{E}(\theta)$ равны θ. Следовательно, выборочное среднее $\overline{X}$ есть оценка по методу моментов параметра θ любого из этих распределений. Легко видеть, что эта оценка состоятельна и несмещена. Точно так же у нормального распределения $\mathcal{N}(\mu,\,\sigma^2)$ параметр μ означает среднее значение, а σ^2 – дисперсию этого распределения. Следовательно, выборочное среднее $\overline{X}$ и выборочная дисперсия S^2 есть состоятельные оценки соответствующих компонент μ и σ^2 параметрического вектора $\theta = (\mu,\,\sigma^2)$. Исправляя смещение оценки S^2 компоненты σ^2, получаем несмещенную оценку θ. Замечательно то, что все оценки являются достаточными статистиками, и это обстоятельство, как будет видно в дальнейшем, предопределяет их оптимальные свойства. Распределение оценки $\overline{X}$ легко получить, используя теоремы сложения для распределений B, P, E и $\mathcal{N}$, распределение же S^2 при выборе из нормального распределения мы найдем несколько позже.

Рассмотрим теперь примеры, в которых приходится решать систему уравнений, и найденные оценки по методу моментов не являются функциями достаточных статистик.

Пример 3.1. *Оценка параметров биномиального распределения* $\mathrm{B}(m,\,p)$. Проблема состоит в оценке обеих компонент m и p двумерного параметра $\theta = (m,\,p)$. Из курса теории вероятностей нам известно, что среднее значение биномиального распределения равно mp, а дисперсия – $mp(1-p)$. Приравнивая эти теоретические моменты их выборочным аналогам, получаем систему для определения оценок по методу моментов: $mp = \overline{X}$, $mp(1-p) = S^2$. Разделив второе уравнение на первое, находим оценку $\hat{p}_n = (\overline{X} - S^2)/\overline{X}$ параметра p, после чего, обращаясь к первому уравнению, находим оценку $\hat{m}_n = \overline{X}^2/(\overline{X} - S^2)$ параметра m. Легко показать, что эти оценки обладают свойством состоятельности (общий метод доказательства таких утверждений смотрите в приведенной ниже теореме 3.1), но при малых n велика вероятность получить отрицательные значения оценок, оценка $\hat{p}_n$ может принять значения, большие единицы, оценка параметра m, как правило, не будет целым числом, наконец, можно показать, что оценка $p_n^* = \overline{X}$ параметра p будет обладать меньшим квадратичным риском, чем оценка $\hat{p}_n$. Все это, конечно, печально, однако другие методы, приводящие к более точным оценкам, обладают значительными вычислительными трудностями.

Пример 3.2. *Оценка параметров гамма-распределения* $\mathrm{G}(\lambda,\, a)$. У этого двухпараметрического распределения среднее равно λa, а дисперсия – λa^2. Решение системы уравнений $\lambda a = \overline{X},\ \lambda a^2 = S^2$ дает оценки $\hat{a}_n = S^2/\overline{X},\ \hat{\lambda}_n = \overline{X}^2/S^2$, которые, как и в предыдущем примере, не являются функциями достаточной статистики $\left(\sum_1^n X_k,\, \prod_1^n X_k\right)$, и как показывают не совсем простые вычисления, их риски далеки от возможного минимума. Тем не менее, очевидная вычислительная простота оценок параметров гамма-распределения по методу моментов обеспечивает их популярность в практических применениях.

Изучим теперь асимптотические свойства оценок по методу моментов – установим условия их состоятельности и исследуем поведение их распределений при больших объемах выборок. Для простоты мы ограничимся случаем одномерного параметра θ, оценка которого определяется решением уравнения $\mu(\theta) = \mathbf{E}_\theta X = \overline{X}$, и предположим, что это уравнение имеет единственное решение $\hat{\theta}_n = h(\overline{X})$. Понятно, что $h(\cdot) = \mu^{-1}(\cdot)$, так что $h(\mu(\theta)) \equiv \theta$. О возможности распространения наших результатов на случай векторного θ мы поговорим отдельно.

Теорема 3.1. *Если наблюдаемая случайная величина X имеет конечное среднее значение $\mu = \mu(\theta)$ и функция $h(\cdot)$ непрерывна в области значений выборочного среднего $\overline{X}$, то $\hat{\theta}_n = h(\overline{X})$ является состоятельной оценкой параметра θ по методу моментов.*

Доказательство.. Мы воспользуемся формулой (1) в определении состоятельности оценки и покажем, что для любых $\varepsilon > 0$ и $\alpha > 0$ существует такое $N(\varepsilon,\, \alpha)$, что для всех $n > N(\varepsilon,\, \alpha)$ вероятность

$$P_\theta\left(\left|h(\overline{X}) - \theta\right| \le \varepsilon\right) > 1 - \alpha. \tag{2}$$

Поскольку $h(\mu) = h(\mu(\theta)) = \theta$ и $\overline{X} \underset{P}{\to} \mu$, то нам достаточно показать, что свойство (или определение) непрерывности функции: $h(x) \to h(\mu)$ при $x \to \mu$, остается справедливым при замене обычной сходимости " $\to$ " на сходимость по вероятности " $\underset{P}{\to}$ ", то есть $\overline{X} \underset{P}{\to} \mu$ влечет $h(\overline{X}) \underset{P}{\to} h(\mu)$. Это почти очевидно, поскольку событие, состоящее в попадании в окрестность нуля случайной величины $|\overline{X} - \mu|$, влечет аналогичное событие для случайной величины $|h(\overline{X}) - h(\mu)|$, но все же проведем строгое доказательство на языке "$\varepsilon - \delta$".

Так как $\overline{X} \underset{P}{\to} \mu(\theta)$, а $h(\cdot)$ – непрерывная функция, то найдутся такие $\delta = \delta(\varepsilon, \alpha)$ и $N = N(\varepsilon, \alpha)$, что

$$P(\,|\overline{X} - \mu(\theta)| \,<\, \delta\,) \,>\, 1 - \alpha \tag{3}$$

для всех $n > N$ и *событие* $|\overline{X} - \mu(\theta)| \,<\, \delta$ *повлечет событие* $|h(\overline{X}) - h(\mu(\theta)| = |h(\overline{X}) - \theta| \leq \varepsilon$. В силу этого неравенство (2) становится следствием неравенства (3). Теорема доказана.

Анализ доказательства показывает, что теорема состоятельности остается справедливой в случае векторного параметра θ, если воспользоваться определением непрерывности векторной функции от векторного аргумента, связав его с расстояниями в эвклидовых пространствах значений функции и ее аргумента.

Обратимся теперь к асимптотическому анализу распределения оценки $\hat{\theta}_n = h(\overline{X})$ при $n \to \infty$. Понятно, что в данном случае употребление термина "оценка" применительно к функции $h(\overline{X})$ ничего особенно не добавляет – речь идет просто об асимптотическом распределении статистики, имеющей вид функции от выборочного среднего.

Теорема 3.2. *Если $X^{(n)}$ – случайная выборка из распределения с конечными средним значением μ и дисперсией σ^2, а функция $h(x)$ обладает ограниченной второй производной $h''(x)$ в некоторой окрестности точки $x = \mu$, то*

$$\lim_{n\to\infty} P\left(\sqrt{n}\left(h(\overline{X}) - h(\mu)\right) \,<\, x\right) = \Phi\left(\frac{x}{\sigma|\,h'(\mu)\,|}\right). \tag{4}$$

где $\Phi(\cdot)$ – функция распределения стандартного нормального закона.

Д о к а з а т е л ь с т в о.. Понятно, что мы должны воспользоваться центральной предельной теоремой (§14 курса ТВ) применительно к статистике $\overline{X} = \sum_1^n X_k$:

$$\lim_{n\to\infty} P\left(\sqrt{n}\left(\overline{X} - \mu\right) \,<\, x\right) = \Phi\left(\frac{x}{\sigma}\right). \tag{5}$$

Стандартный прием использования этой теоремы при асимптотическом анализе функций от сумм независимых, одинаково распределенных случайных величин состоит в “линеризации” таких функций с помощью формулы Тейлора. В нашем случае мы разлагаем функцию $h(\cdot)$ в окрестности точки $\overline{X} = \mu$, используя только два члена разложения:

$$h(\overline{X}) = h(\mu) + (\overline{X} - \mu)h'(\mu) + \frac{(\overline{X} - \mu)^2}{2!}h''\left(\mu + \lambda(\overline{X} - \mu)\right),$$

где $0 \le \lambda \le 1$.

Перепишем это разложение в виде

$$\sqrt{n}\left(h(\overline{X}) - h(\mu)\right) = \sqrt{n}(\overline{X} - \mu)h'(\mu) + \frac{\sqrt{n}(\overline{X} - \mu)^2}{2!}h''\left(\mu + \lambda(\overline{X} - \mu)\right),$$

представив тем самым случайную величину $\sqrt{n}\left(h(\overline{X}) - h(\mu)\right)$ (см. формулу (4)) в виде суммы двух случайных величин, первая из которых в силу формулы (5) имеет предельное нормальное распределение, указанное в правой части (4), а вторая сходится по вероятности к нулю. Действительно, $h''\left(\mu + \lambda(\overline{X} - \mu)\right)$ по условию теоремы ограничено с вероятностью сколь угодно близкой к единице, начиная с некоторого n. В силу неравенства Чебышева (предложение 6.2 курса ТВ) для любого $\varepsilon > 0$ вероятность

$$P\left((\overline{X} - \mu)^2 \,>\, \varepsilon\right) \le \frac{\mathbf{E}\sqrt{n}(\overline{X} - \mu)^2}{\varepsilon} = \frac{\sqrt{n}\mathbf{D}\overline{X}}{\varepsilon} = \frac{\sigma^2}{\varepsilon\sqrt{n}} \to 0,$$

и поэтому стоящий перед $h''/2!$ множитель, а с ним и все второе слагаемое, сходятся по вероятности к нулю. Доказательство теоремы завершается ссылкой на предложение 11.1 курса ТВ: если одна последовательность случайных величин имеет невырожденное предельное распределение F, а вторая сходится по вероятности к нулю, то предельное распределение суммы этих последовательностей совпадает с F.

Итак, если $h(\overline{X})$ – оценка $\hat{\theta}_n$ параметра θ по методу моментов, то формула (4) теоремы 3.2 принимает вид

$$\lim_{n\to\infty} P\left(\sqrt{n}\left(\hat{\theta}_n - \theta\right)\right) \, < \, x \Big) = \Phi\left(\frac{x}{\sigma|\, h'(\mu)\,|}\right).$$

Приведем пример на использование аппроксимации распределения функции от выборочного среднего на практике.

П р и м е р 3.3. *Оценка надежности изделия с показательным распределением долговечности.* При выпуске изделий обычно указывается их гарантийный срок службы t_0; отказ изделия до истечения срока t_0 чреват для поставщика расходами на ремонт или замену изделия. Чтобы планировать расходы на такого рода рекламации со стороны потребителя, поставщик должен знать надежность выпускаемых изделий: $H(t_0) = P(X > t_0)$. Величина $H(t_0)$ указывает среднюю долю среди выпущенных изделий, которые могут отказать за гарантийное время службы t_0. Чтобы оценить $H(t_0)$ проводятся испытания n изделий, и пусть $x_1, \ldots, x_n$ – наработки на отказ испытуемых изделий, трактуемые как реализации случайной выборки $X^{(n)}$ из распределения $F(\,\cdot\,|\,\theta)$, известного с точностью до значения параметра θ. Наконец, пусть нам известно, что долговечность изделий подчиняется закону “отсутствия последействия”, в силу чего $F(x\,|\,\theta) = 1 - \exp\{-x/\theta\}$ – показательное распределение. В таком случае проблема состоит в оценке параметрической функции $h(\theta) = \exp\{-t_0/\theta\}$.

Так как $\theta = \mathbf{E}X$ – средняя наработка на отказ, то естественно оценить θ посредством статистики $\overline{X}$ (работает метод моментов), а за оценку надежности взять статистику $h(\overline{X}) = \exp\{-t_0/\overline{X}\}$. Поскольку наибольшую, с экономической точки зрения, опасность представляет завышение надежности, то нас в первую очередь должна интересовать частота грубых превышений, например, на некоторую заданную величину ε. Если положить $\Delta = \varepsilon\sqrt{n}$, то вероятность, указывающая частоту грубых превышений, записывается в виде $R(\theta;\, h(\overline{X})) = P\left(\sqrt{n}\left(h(\overline{X}) - h(\theta)\right) \, > \, \Delta\right)$, что позволяет нам непосредственно использовать аппроксимацию (4) при испытаниях достаточно большого числа изделий (насколько большого, это – отдельный вопрос, решить который можно, например, моделируя выборки из показательного распределения с помощью метода Монте-Карло).

Имеем $\sigma^2 = \theta^2$, $h'(\theta) = t_0\theta^{-2}\exp\{-t_0/\theta\}$, и формула (4) дает нам следующую аппроксимацию для риска оценки $h(\overline{X})$:

$$R(\theta;\, h(\overline{X})) \approx 1 - \Phi\left(\frac{\Delta\theta}{t_0}\exp\left\{\frac{t_0}{\theta}\right\}\right).$$

Нетрудно показать, что наибольшее значение риска достигается при $\theta = t_0$ и равно $\Phi(\Delta \cdot \mathrm{e})$.

В том случае, когда оценка имеет вид функции от двух и более выборочных моментов, метод асимптотического анализа ее распределения тот же. Например, пусть $\hat{\theta}_n = h(a_1,\, a_2)$, и функция h удовлетворяет условиям, аналогичным требованиям к h в теореме 3.2. Используя формулу Тейлора, представим h в окрестности точки $(\alpha_1,\, \alpha_2)$ в следующем виде:

$$h(a_1,\, a_2) = h(\alpha_1,\, \alpha_2) + (a_1 - \alpha_1)h_1'(\alpha_1,\, \alpha_2) +$$

$$(a_2 - \alpha_2)h_2'(\alpha_1,\, \alpha_2) + O_p(|\, a - \alpha\,|^2),$$

где $a = (a_1,\, a_2)$, $\alpha = (\alpha_1,\, \alpha_2)$. Легко проверяется, что $\sqrt{n}|\, a - \alpha\,|^2 \underset{P}{\to} 0$, так что случайная величина $\sqrt{n}\,(h(a_1,\, a_2) - h(\alpha_1,\, \alpha_2))$ асимптотически нормальна с параметрами, которые выражаются через первые четыре момента наблюдаемой случайной величины X.