Дорогие студенты, вы зашли в так называемую “Виртуальную аудиторию,” в которой к каждому понедельнику я буду заносить материал новых лекций. Если у вас возникают какие-либо вопросы, то вы можете присылать их мне по электронной почте

volodinstudent@gmail.com

На ваши вопросы я буду отвечать вам reply’ем. В конце текста каждой лекции будут предложены задания, которые вам следует выполнить и выслать мне по почте (или, по крайней мере, сообщать, что вы “на проводе”). О том, как их выполнить я могу вам рассказать при общении в Teams’е.

С надеждой на скорое закрытие карантина ваш преподаватель Игорь Николаевич.

Лекция 8

Статистическая идентификация гауссовских моделей: точечная и интервальная оценка параметров модели

Сегодняшняя лекция посвящена задачам оценки параметров нормального распределения. Эта тема хорошо известна вам, и я решил рассказать вам кое-какие известные факты, которые будут нужны вам при сдачи экзаменов.

Итак, рассмотрим нормальное распределение $\mathcal{N}(\mu,\,\sigma^2)$. Это распределение обладает двумерным параметром $\theta=(\mu,\,\sigma)$ с областью значений (параметрическим пространством) $\Theta=\mathrm{R}\times\mathrm{R}_+$. Функции плотности наблюдаемой случайной величины ξ и случайной выборки $X^{(n)}$ определяются соответственно как

$$f(x\,|\,\theta)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

и

$$f_n(x^{(n)}\,|\,\theta)=\frac{1}{(2\pi)^{n/2}\sigma^n}\exp\left\{-\frac{1}{2\sigma^2}\sum_1^n(x_k-\mu)^2\right\}=$$

$$=\frac{1}{(2\pi)^{n/2}\sigma^n}\exp\left\{-\frac{1}{2\sigma^2}\left(\sum_1^n x_k^2-2\mu\sum_1^n x_k+n\mu^2\right)\right\}.$$

Последнее выражение для плотности $X^{(n)}$ показывает, что двумерная статистика $T=(T_1,\,T_2)$ с

$$T_1=\sum\nolimits_1^n X_k\quad T_2=\sum\nolimits_1^n X_k^2$$

достаточна для статистической структуры нормального распределения. Кроме того, поскольку $T_1=n\overline{X}$ и $T_2=n(S^2+\overline{X}^2)$, то факторизационное равенство Неймана указывает на достаточность статистик $\overline{X}$ и S^2, которые имеют конкретную статистическую интерпретацию и поэтому более удобны для практического использования. Понятно, что это замечание носит общий характер: *любые взаимно однозначные преобразования достаточной статистики наследуют свойство достаточности.*

Отметим также, что в случае известного (фиксированного) σ статистическая структура имеет параметрическое пространство, совпадающее с областью значений параметра μ, и достаточной статистикой будет выборочное среднее $\overline{X}$. Аналогичное утверждение имеет место для статистики S^2 при фиксированном μ.

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 предопределяет их оптимальные свойства. Распределение оценки $\overline{X}$ легко получить, используя теоремы сложения для распределений B, P, E и $\mathcal{N}$,

Рассмотрим теперь более точный метод оценки неизвестного значения параметра. Он превосходит метод моментов и при наличии достаточных статистик дает оптимальные оценки с точки зрения квадратичного риска. Более того, при выполнении определенных условий регулярности этот метод приводит к асимптотически $(n \to \infty)$ оптимальным оценкам для широкого класса вероятностных моделей и практически при любых функциях потерь.

Для случайной величины $\xi \sim \mathcal{N}(\mu, \sigma^2)$ функция плотности выборки

$$f_n(x^{(n)} \,|\, \theta) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left\{ -\frac{1}{2\sigma^2} \sum\nolimits_1^n (x_k - \mu)^2 \right\},$$

при каждом фиксированном значении двумерного параметра $\theta = (\mu,\, \sigma)$ служит функцией правдоподобия для выборочных данных $x^{(n)} = (x_1, \ldots, x_n)$. Точки достижения максимума функции $f_n(X^{(n)} \,|\, \mu, \sigma)$ по переменным $\mu \in \mathrm{R}$ и $\sigma \in \mathrm{R}_+$ доставляют оценки этих параметров. Естественно, логарифм этой функции имеет те же точки экстремума, что и сама функция, но логарифмирование упрощает выкладки, поэтому ищем максимум функции

$$\mathcal{L}(\theta \,|\, X^{(n)}) = \ln f_n(X^{(n)} \,|\, \theta) = -\frac{n}{2} \ln 2\pi - n \ln \sigma - \frac{1}{2\sigma^2} \sum\nolimits_1^n (X_k - \mu)^2.$$

Составляем уравнения, определяющие точки экстремума:

$$\frac{\partial \mathcal{L}}{\partial \mu} = \frac{1}{\sigma^2} \sum\nolimits_1^n (X_k - \mu) = 0,$$

$$\frac{\partial \mathcal{L}}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum\nolimits_1^n (X_k - \mu)^2 = 0.$$

Из первого уравнения сразу находим оценку параметра $\mu: \quad \hat{\mu}_n = \overline{X}$. Подставляя $\overline{X}$ вместо μ во второе уравнение, находим оценку $\sigma: \quad \hat{\sigma}_n = S$ (выборочное стандартное отклонение). Очевидно, $(\overline{X},\, S)$ – точка максимума. Таким образом, метод максимального правдоподобия приводит к тем же оценкам $\overline{X}$ и S^2 параметров μ и σ^2, что и метод моментов.

Рассмотри еще один практически важный пример по использованию метода максимального правдоподобия в рамках нормального распределения наблюдаемой случайной величины. Это, так называемая, *оценка параметров структурированного среднего при нормальном распределении отклика*. Данная задача весьма часто возникает при калибровке шкалы прибора. Две переменные x и y связаны линейным соотношением $y = a + bx$, и для градуировки значений y на шкале прибора необходимо знать значения параметров a и b этой зависимости. Однако для каждого стандартного фиксированного значения x прибор замеряет значение y с ошибкой, так что замеры происходят в рамках вероятностной модели $Y = a + bx + \xi$, где ошибка измерения (случайная величина) ξ имеет нормальное распределение с нулевым средним и некоторой дисперсией σ^2, значение которой, как правило, также неизвестно. Случайная величина Y обычно называется *откликом* на значение *регрессора* x; ее распределение при фиксированном x очевидно нормально $(a + bx,\, \sigma^2)$.

Для оценки a и b производится n измерений отклика $y_1, \ldots, y_n$ при некоторых фиксированных значениях $x_1, \ldots, x_n$ регрессора x, оптимальный выбор которых, обеспечивающий наибольшую точность и надежность калибровки, составляет самостоятельную задачу особой области математической статистики – *планирование регрессионных экспериментов*. Мы будем предполагать, что значения $x_1, \ldots, x_n$ априори фиксированы. В таком случае значения $y_1, \ldots, y_n$ представляют реализации n независимых случайных величин $Y_1, \ldots, Y_n$, и $Y_k \sim \mathcal{N}(a + bx_k,\, \sigma^2),\ k = 1, \ldots, n$. Совместная функция плотности $Y_1, \ldots, Y_n$ равна

$$f_n(y^{(n)} \,|\, a, b, \sigma) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left\{-\frac{1}{2\sigma^2} \sum\nolimits_1^n (y_k - a - bx_k)^2)\right\},$$

так что логарифмическая функция правдоподобия, необходимая для оценки параметров a, b и σ методом максимального правдоподобия имеет вид

$$\mathcal{L}(a,\, b,\, \sigma \,|\, Y^{(n)}) = -\frac{n}{2} \ln 2\pi - n \ln \sigma - \frac{1}{2\sigma^2} \sum_1^n (Y_k - a - bx_k)^2.$$

Вычисляя производные этой функции по переменным a, b и σ, получаем уравнения правдоподобия

$$\sum_1^n (Y_k - a - bx_k) = 0,$$

$$\sum_1^n x_k (Y_k - a - bx_k) = 0,$$

$$n\sigma^2 - \sum_1^n (Y_k - a - bx_k)^2 = 0.$$

Конечно, это очень простая система уравнений, решение которой не может вызывать каких-либо затруднений, и мы сразу пишем оценки максимального правдоподобия

$$\hat{a}_n = \overline{Y} - \frac{m_{xY}}{s_x^2} \overline{x}, \quad \hat{b}_n = \frac{m_{xY}}{s_x^2}, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_1^n \left(Y_k - \hat{a}_n - \hat{b}_n x_k\right)^2,$$

где

$$\overline{x} = \frac{1}{n} \sum_1^n x_k,\ \overline{Y} = \frac{1}{n} \sum_1^n Y_k, \quad s_x = \frac{1}{n} \sum_1^n (x_k - \overline{x})^2,\ S_Y = \frac{1}{n} \sum_1^n (Y_k - \overline{Y})^2,$$

$$m_{xY} = \frac{1}{n} \sum_1^n (x_k - \overline{x})(Y_k - \overline{Y}).$$

Легко видеть, что оценки по методу максимального правдоподобия параметров a и b совпадают с их оценками по *методу наименьших квадратов*. В этом методе "выравнивания" экспериментальных данных оценки ищутся из условия минимизации суммы квадратов *невязок*: $\sum_1^n (Y_k - a - bx_k)^2$, причем под невязкой понимается разность между откликом Y и его "теоретическим" средним значением $a + bx$.

Мы рассмотрели несколько методов построения *точечных* оценок для параметров, значения которых определяют распределение наблюдаемой случайной величины. Был получен ряд утверждений о распределении таких оценок, что позволяет судить о надежности оценки при заданной точности, то есть вычислять вероятности событий вида $|\,\hat{\theta}_n(X^{(n)}) - \theta\,| \leqslant \Delta$ при каждом фиксированном значении параметра θ. Поскольку именно значение θ нам неизвестно, то такого рода вычисления зачастую лишены практического смысла – слишком велик размах в надежности оценки $\hat{\theta}_n$ при различных θ, даже в случае когда мы располагаем некоторой априорной информацией о возможной области значений этого параметра. Поэтому в ряде практических ситуаций пытаются решать обратную задачу: для фиксированной надежности, скажем, $1 - \alpha$, где α мало, указать некоторую область значений θ, зависящую, естественно, от выборки $X^{(n)}$, которая с вероятностью, не меньшей $1 - \alpha$, накрывает истинное, неизвестное нам значение θ, причем такое надежностное утверждение должно выполняться при любых $\theta \in \Theta$. В таком случае по размерам области, которые определяются выборочными значениями $x^{(n)}$, можно судить о точности такой *интервальной* оценки.

Определение 6.1. Подмножество $\Delta_n = \Delta_n(X^{(n)})$ параметрического пространства Θ называется $(1 - \alpha)$-*доверительной областью*, если

$$P_\theta\left(\Delta_n(X^{(n)}) \ni \theta\right) \geqslant 1 - \alpha, \tag{1}$$

каково бы ни было значение $\theta \in \Theta$. Заданное (фиксированное) значение $1 - \alpha$ называется *доверительным уровнем*, а наименьшее значение левой части неравенства (1) по всем $\theta \in \Theta$ – *доверительным коэффициентом*. В случае $\Theta \subseteq \mathrm{R}$ доверительная область вида $\Delta_n = \left(\underline{\theta}_n(X^{(n)}); \overline{\theta}_n(X^{(n)})\right)$ называется *доверительным интервалом*, в котором различаются *нижний* $\underline{\theta}_n$ и *верхний* $\overline{\theta}_n$ *доверительные пределы*. Доверительные интервалы вида $(\underline{\theta}_n; \infty)$ и $(-\infty; \overline{\theta}_n)$ называются соответственно *нижней* и *верхней доверительными границами*.

Естественно, конфигурация доверительной области выбирается статистиком, сообразуясь с ее геометрической наглядностью и, главное, возможностью гарантировать доверительную вероятность. В случае скалярного параметра доверительная область обычно выбирается в виде интервала, причем в ряде случаев, например, при оценке надежности или вероятности нежелательного события, в виде одностороннего интервала. В случае многомерного параметра обычно строятся доверительные эллипсоиды или параллелепипеды.

Следует обратить особое внимание на правильную формулировку доверительного утверждения, которая подчеркивается в неравенстве (1) записью $\Delta_n(X^{(n)}) \ni \theta$ вместо обычного $\theta \in \Delta_n(X^{(n)})$. Говорить, что значение параметра θ с вероятностью, не меньшей $1 - \alpha$, принадлежит области Δ_n, значит сознательно вводить трудящихся на ниве прикладной статистики в заблуждение. Дело в том, что значение параметра θ в данной вероятностной модели не является случайной величиной, это постоянная, свойственная исследуемому объекту, а постоянная принадлежит какой-либо области только с вероятностью единица или ноль. Вся случайность заключена в самой доверительной области $\Delta_n(X^{(n)})$, и поэтому правильное доверительное утверждение гласит: *область* $\Delta_n(X^{(n)})$ *с вероятностью, не меньшей* $1-\alpha$, *накрывает истинное (неизвестное) значение* θ.

Рассмотрим методы построения доверительных интервалов для оценки среднего значения нормального распределения. Начнем с более простого примера.

1^0. Доверительный интервал для среднего значения нормального распределения при известной дисперсии. Пусть $X^{(n)}$ – случайная выборка из нормального $(\mu,\, \sigma^2)$ распределения, причем значение параметра σ известно, так что в качестве неизвестного параметра θ выступает μ. Наша задача состоит в построении такого интервала $\left(\underline{\mu}_{\,n}(X^{(n)}),\, \overline{\mu}_n(X^{(n)})\right)$, что $P_\mu\left(\underline{\mu}_{\,n} \leqslant \mu \leqslant \overline{\mu}_n\right) \geqslant 1-\alpha$, при любом значении $\mu \in \mathrm{R}$.

В этом примере оценкой μ служит выборочное среднее $\overline{X}$ – несмещенная оценка μ с минимальным квадратичным риском. Эта линейная оценка обладает замечательным свойством инвариантности: распределение разности $\overline{X}-\mu$ не зависит от μ, и это обстоятельство подсказывает нам путь к построению доверительного интервала. Действительно,

$$P\left(\frac{|\,\overline{X}-\mu\,|}{\sigma}\sqrt{n} \,\leqslant\, \lambda\right) = 2\Phi(\lambda)-1,$$

и если положить λ равным корню уравнения $2\Phi(\lambda)-1 = 1-\alpha$, то есть выбрать $\lambda = \lambda_\alpha = \Phi^{-1}(1-\alpha/2)$, то интервал $(\overline{X}-\lambda_\alpha\sigma/\sqrt{n},\ \overline{X}+\lambda_\alpha\sigma/\sqrt{n})$ будет $(1-\alpha)$-доверительным интервалом для среднего значения μ нормального $(\mu,\, \sigma^2)$ распределения при известной дисперсии σ^2.

Рассмотрим теперь задачу построения доверительного интервала для μ, когда значение дисперсии σ^2 не известно. Для решения этой задачи нам потребуется знать совместное распределение выборочного среднего и выборочной дисперсии.

Теорема 1. *В случае выбора из нормального* $(\mu,\, \sigma^2)$ *распределения статистики* $\overline{X}$ *и* S^2 *независимы,* $\overline{X} \sim \mathcal{N}(\mu,\, \sigma^2/n)$, *а* $nS^2/\sigma^2 \sim \chi^2_{n-1}$ *(имеет хи-квадрат распределение с* $n-1$ *степенью свободы).*

Доверительный интервал для среднего значения нормального распределения при неизвестной дисперсии. В этой задаче мы имеем дело с двусторонними доверительными границами (доверительным интервалом) и в соответствии с выбором опорной функции $|\,\overline{X}-\mu\,|/S$ нам потребуется знание вероятности события вида $|\,\overline{X}-\mu\,|/S \leqslant \lambda$.

В начале XIX века английский математик В.Госсет, писавший под псевдонимом "Стьюдент" (Student), нашел распределение случайной величины $T_\nu = \xi\sqrt{\nu}/\sqrt{\chi^2_\nu}$, где $\xi \sim \mathcal{N}(0,\, 1)$, а χ^2_ν – случайная величина, не зависящая от ξ и распределенная по закону хи-квадрат с ν степенями свободы. Естественно, его исследования были связаны с проблемами статистического вывода о среднем значении μ нормального распределения при неизвестной дисперсии, и Стьюдент искал распределение опорной функции в связи с переходом в записи опорной функции в терминах X_k к Y_k)

$$H = \frac{\overline{X} - \mu}{S}\sqrt{n-1} = \frac{\frac{1}{\sqrt{n}}\sum_1^n Y_k}{\sqrt{\sum_1^n (Y_k - \overline{Y})^2}}\sqrt{n-1},$$

$$Y_k = \frac{X_k - \mu}{\sigma} \sim \mathcal{N}(0,\, 1),\ k = 1, \ldots, n,$$

которая отличается от выбранной нами опорной функции только множителем $\sqrt{n-1}$, и поэтому также может быть использована в построении доверительного интервала для μ при неизвестном σ. То, что распределения T_{n-1} и H совпадают, следует из Теоремы 1: случайная величина в числителе

$$\xi = \sum_1^n Y_k/\sqrt{n} \sim \mathcal{N}(0,\, 1)$$

не зависит от

$$\sum_1^n (Y_k - \overline{Y})^2 \sim \chi^2_{n-1},$$

разделив которую на значение степени свободы $n-1$, получаем

$$\frac{1}{\nu}\chi^2_\nu$$

с $\nu = n - 1$.

Распределение случайной величины T_ν называется *распределением Стьюдента с ν степенями свободы* или *t-распределением.*

Обратимся теперь к проблеме оценки параметров многомерного нормального распределения.

При выборе из m-мерного, $m > 1$, распределения эмпирическое распределение также приписывает массу n^{-1} каждому выборочному (векторному) значению $X_i = (X_{1i}, \ldots, X_{mi}),\ i = 1, \ldots, n$. В соответствии с этим мы можем определить вектор выборочных средних $\overline{X} = (\overline{X}_1, \ldots, \overline{X}_m)$ с компонентами

$$\overline{X}_k = \frac{1}{n}\sum_{i=1}^n X_{ki},\ k = 1, \ldots, m,$$

выборочную ковариационную матрицу $S = \| S_{kj} \|$ с элементами

$$S_{kj} = \frac{1}{n}\sum_{i=1}^n (X_{ki} - \overline{X}_k)(X_{ji} - \overline{X}_j) = \frac{1}{n}\sum_{i=1}^n X_{ki}X_{ji} - \overline{X}_k\overline{X}_j,\ k, j = 1, \ldots, m,$$

и матрицу *выборочных коэффициентов корреляции* $R = \| r_{kj} \|$ с элементами

$$r_{kj} = S_{kj}/\sqrt{S_{kk}S_{jj}},\ k, j = 1, \ldots, m.$$

Смешанные моменты более высоких порядков в многомерном случае обычно не вычисляются.

В качестве примера рассмотрим задачу *оценки параметров двумерного нормального распределения.*

Оценка по методу максимального правдоподобия пяти параметров μ_1, μ_2, σ_1^2, σ_2^2, ρ двумерного нормального распределения с функцией плотности

$$f(x, y \,|\, \theta) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot$$

$$\exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right)\right\}$$

не представляет особой технической сложности. Эти оценки совпадают с оценками по методу моментов и, таким образом, равны выборочным аналогам тех характеристик двумерного нормального распределения, которые соответствуют указанным пяти параметрам:

$$\hat{\mu}_{1,n} = \overline{X},\ \hat{\mu}_{2,n} = \overline{Y},\ \ \hat{\sigma}^2_{1,n} = S_X^2,\ \hat{\sigma}^2_{2,n} = S_Y^2,\ \ \hat{\rho}_n = r.$$

Формулы для вычисления выборочных средних $\overline{X}$ и $\overline{Y}$, выборочных дисперсий S_1^2 и S_2^2, а также выборочного коэффициента корреляции r приведены выше.

Полученные оценки часто используются для оценки параметров линейного прогноза $Y = a + bX$ значений случайной величины Y по результатам наблюдений X. В случае нормального распределения линейный прогноз обладает свойством оптимальности с точки зрения малости средней квадратичной ошибки и совпадает с кривой средней квадратичной регрессии

$$y = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1).$$

Однако формальная подгонка прогностической кривой с помощью прямой линии используется и вне рамок нормальной модели, и в этом случае оценки

$$\hat{a}_n = \overline{Y} - r\frac{S_2}{S_1}\overline{X}, \quad \hat{b}_n = r\frac{S_2}{S_1}$$

совпадают с оценками по *методу наименьших квадратов*: минимизируется сумма квадратов невязок

$$\sum_1^n (Y_k - a - bX_k)^2.$$

Хотя оценки в этом примере и примере с оценкой структурированного среднего имеют одинаковый вид, но решаемые в них статистические проблемы весьма различны: в первом примере оценивались параметры некоторой функциональной зависимости с ошибками в наблюдениях отклика, в то время как в последнем примере решается задача выявления корреляционной связи и использования этой связи для прогноза.

Контрольные вопросы.

1.Оценить параметр θ нормального распределения, у которого среднее совпадает со стандартным распределением, то есть $\xi \sim \mathcal{N}(\theta,\, \theta^2)$

2. Используя опорную функцию $\sigma/|\,\xi\,|$ постройте верхнюю $1-\alpha$-доверительную границу для σ по одному наблюдению $\xi \sim \mathcal{N}(\mu,\, \sigma^2)$ при неизвестном значении μ.