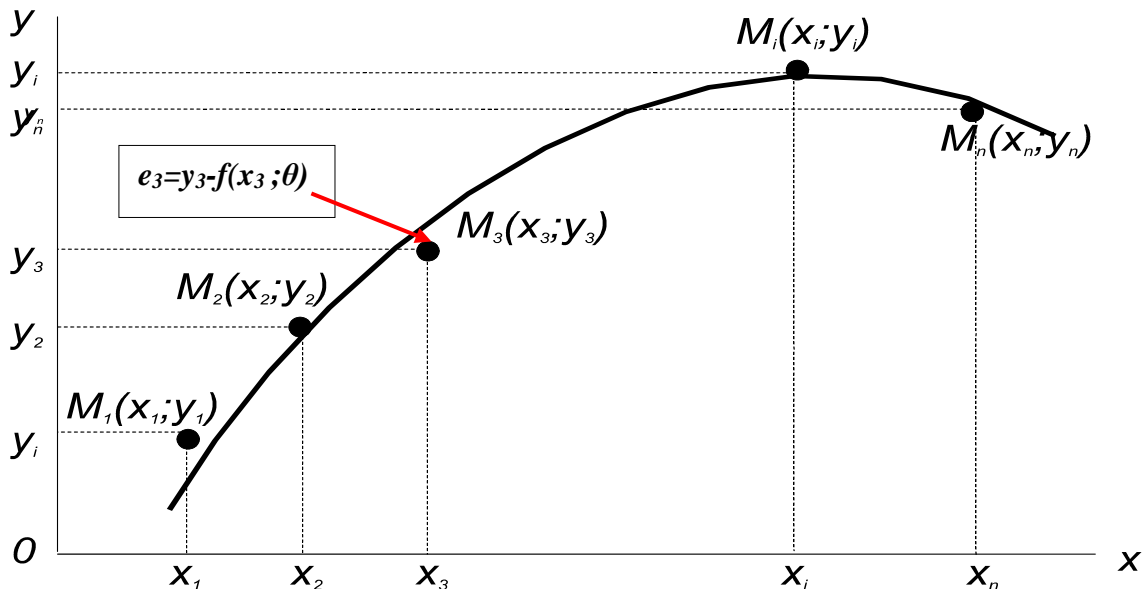


СМОДЗУ, Лекция 5.

Методы определения зависимостей.

Задача ставится следующим образом: даны n пар значений (x_i, y_i) $i=1, \dots, n$, где x_i – моменты времени или значения какого-либо другого



аргумента, при которых происходят измерения интересующей исследователя величины y_i . Измерения выполняются независимо и с некоторой ошибкой σ , величина которой также подлежит оценке. Требуется подогнать наилучшим образом к этим точкам некую кривую, т.е. определить вид функциональной зависимости y от x . Параметрическое семейство функций $y=f(x; \bar{\vartheta})$, где $\bar{\vartheta}$ – вектор параметров, может быть известно и тогда оценке подлежит именно $\bar{\vartheta}$.

Однако задача может ставиться так, что **вид зависимости неизвестен**, и тогда требуется сравнить результаты подгонки нескольких семейств функций и выбрать вариант, наилучший с точки

зрения некоего статистического критерия качества подгонки.

Из рисунка видно, что для каждого значения x_i экспериментальное y_i и расчетное y_i^p значения различаются на некоторую величину $e_i = y_i - f(x_i; \bar{\theta})$, называемую **невязкой** в i -ой точке.

3.1. Метод максимального правдоподобия.

В предположении, что все невязки являются независимыми случайными величинами с нулевыми средними и одинаковой функцией плотности распределения (ф.п.р.) $\varphi(\mathbf{e})$, вероятность появления конкретной выборки $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ определяется функцией правдоподобия

$$L(\bar{\theta}) = \prod_{i=1}^n \varphi(e_i) \quad (3.1)$$

Согласно принципу максимального правдоподобия Р.Фишера наилучшей - наиболее правдоподобной - оценкой параметров будет тот набор параметров $\bar{\theta} = \theta_1, \theta_2, \dots, \theta_r$, при котором $L(\bar{\theta})$ примет свое максимальное значение. Поскольку и сама функция правдоподобия, и ее логарифм принимают максимальное значение в одной и той же точке, то можно перейти к логарифмической функции правдоподобия $l(\bar{\theta}) = \ln L(\bar{\theta})$, что позволяет в большинстве случаев упростить процедуру дифференцирования при поиске максимума. Соответствующее уравнение правдоподобия для поиска максимального значения $\bar{\theta}$ записывается так:

$$\frac{\partial \ln L}{\partial \theta_k} = \sum_{i=1}^n \ln \varphi(e_i) = 0, k = 1, 2, \dots, r. \quad (3.2)$$

Решение этого уравнения и дает нам ММП-оценки $\hat{\theta}$, которые, как доказано, обладают свойствами состоятельности, асимптотической нормальности и эффективности.

Предположение о нормальности ф.п.р. в (3.1), т.е.

$$\varphi(e_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{e_i^2}{2\sigma^2}} \quad (3.3)$$

является весьма важным, так как в этом случае метод максимального правдоподобия (ММП) превращается в **метод наименьших квадратов (МНК)** с сохранением всех полезных свойств получаемых оценок.

3.1. Метод наименьших квадратов.

В самом деле для плотности распределения вида (3.3) логарифмическая функция правдоподобия оказывается равной величине

$$\ln L = \sum_{i=1}^n \ln \varphi(e_i) = -\frac{1}{2\sigma^2} \sum_{i=1}^n e_i^2 + const ,$$

которая имеет свой максимум там же, где свой минимум имеет функционал МНК:

$$S = \sum_{i=1}^n e_i^2 . \quad (3.4)$$

Уравнение правдоподобия превращается для **S** в более простое, так называемое нормальное уравнение

$$\sum_{i=1}^n e_i e_i' = 0 , \quad (3.5)$$

где штрих означает дифференцирование по параметру. В случаях, когда измерения y_i имеют разную точность σ_i невязки e_i должны входить в

формулы (3.4)-(3.5) с разными весами $w_i = \sigma_i^{-2}$, т.е. уравнение (3.5), например, заменяется на

$$\sum_{i=1}^n w_i e_i e_i' = 0 \quad (3.6)$$

Пример применения МНК для случая подгонки измеренных точек прямой линией.

Будем искать приближающую функцию в виде: $y = f(x, k, b) = kx + b$. В этом случае $e_i = y_i - kx_i - b$ функционал (3.4) переписывается в виде:

$$S(k, b) = \sum_{i=1}^n (y_i - (kx_i + b))^2.$$

Задача сводится к отысканию минимума **$S(k, b)$** . Используем необходимое условие экстремума:

$$\frac{\partial S(k, b)}{\partial k} = 0; \quad \frac{\partial S(k, b)}{\partial b} = 0,$$

т.е.

$$\frac{\partial S}{\partial k} = \frac{\partial \sum_{i=1}^n (y_i - (kx_i + b))^2}{\partial k} = 0; \quad \frac{\partial S}{\partial b} = \frac{\partial \sum_{i=1}^n (y_i - (kx_i + b))^2}{\partial b} = 0.$$

Решив систему двух уравнений с двумя неизвестными относительно параметров **k** и **b** , получим оценки искомых параметров:

$$\hat{k} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}; \quad \hat{b} = \frac{1}{n} \left(\sum_{i=1}^n y_i - k \sum_{i=1}^n x_i \right).$$

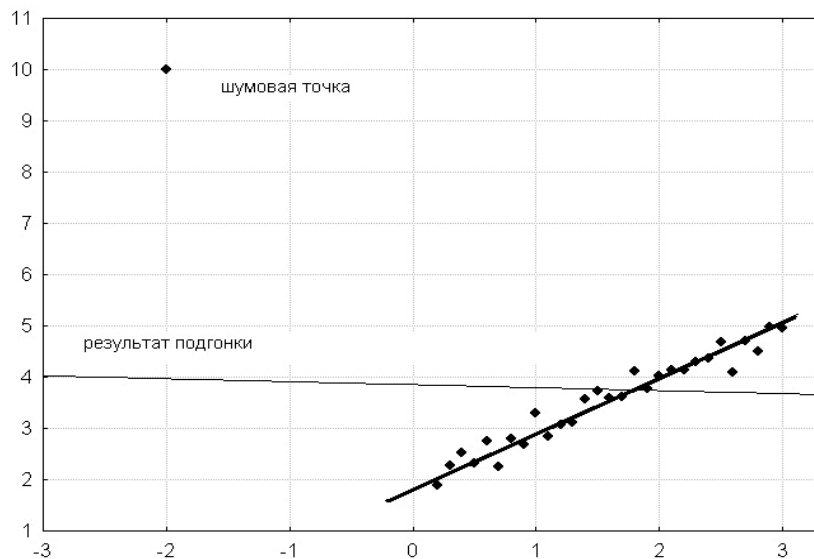
Рассчитав значение **S** в минимуме $S_{\min} = \sum_{i=1}^n (y_i - (\hat{k}x_i + \hat{b}))^2$,

получим величину среднеквадратичной ошибки

рассматриваемого приближения $\hat{\sigma} = \sqrt{\frac{S_{\min}}{n-2}}$,

где число степеней свободы S_{min} получено уменьшением объема выборки на число параметров.

3.4. Робастные методы подгонки



К сожалению, ключевое предположение о нормальности невязок в реальной жизни нарушается очень часто благодаря засорению измерений шумовыми или фоновыми измерениями, как например, показано на рисунке, где всего одна шумовая точка полностью испортила результаты подгонки. Квадратичность функционала S в (3.4) ведет к тому, что далеко отстоящие точки могут дать неоправданно большой вклад в функционал и привести к значительной потере точности оценок параметров.

Чтобы избежать этого следует учитывать измерения только из непосредственной окрестности подгоняемой функции, придавая остальным меньшие значения или вообще пренебрегая ими. Такую идею можно реализовать придавая каждому измерению специальный вес, значение которого убывает с ростом невязки e_i , т.е. расстояния до подгоняемой

кривой. Этот подход, называемый **робастным** (robust (англ.) - крепкий, здоровый, в статистике - не чувствительный к шумам) был предложен П.Хьюбером.

Выберем весовую функции, оптимальную при подгонке кривой на фоне равномерного засорения. Будем по-прежнему предполагать, что измерения производятся с ошибками, распределенными нормальному закону (3.3) со стандартным отклонением σ , а засоряющие точки распределены равномерно на гораздо более широком интервале Δ ($\sigma \ll \Delta$). Предположим также, что отношение сигнал-шум равно $(1-\varepsilon)/\varepsilon$, т.е. ε - это параметр засорения.

Для описания засоренного распределения воспользуемся моделью "больших ошибок" Тьюки, т.е. примем плотность распределения равной

$$f_{\varepsilon}(e) = (1 - \varepsilon)\phi(e) + \varepsilon h(e),$$

где $\phi(e)$ - гауссово распределение (3.3), а $h = 1/\Delta$ - плотность равномерного распределения.

Составим логарифмическую функцию правдоподобия для такого распределения, рассматривая по-прежнему для простоты однопараметрическую зависимость с параметром p

$$L(p) = \ln \prod_i f(e_i) = \sum_i \ln \left(\frac{1 - \varepsilon}{\sqrt{2\pi}\sigma} e^{-\frac{e_i^2}{2\sigma^2}} + \frac{\varepsilon}{\Delta} \right).$$

Приравнивая к нулю ее производную по параметру p , мы получим уравнение правдоподобия

$$\frac{\partial \ln L}{\partial p} = \sum_i \frac{\frac{1 - \varepsilon}{\sqrt{2\pi}\sigma} e^{-\frac{e_i^2}{2\sigma^2}} e_i \frac{\partial e_i}{\partial p}}{\frac{1 - \varepsilon}{\sqrt{2\pi}\sigma} e^{-\frac{e_i^2}{2\sigma^2}} + \frac{\varepsilon}{\Delta}} = 0,$$

которое может быть переписано в виде, аналогичном (3.6):

$$\frac{\partial \ln L}{\partial p} = \sum_i w(e_i) e_i \frac{\partial e_i}{\partial p} = 0, \quad (3.7)$$

где обозначено

$$w(e) = \frac{\frac{1-\varepsilon}{\sqrt{2\pi}\sigma} e^{-\frac{e^2}{2\sigma^2}}}{\frac{1-\varepsilon}{\sqrt{2\pi}\sigma} e^{-\frac{e^2}{2\sigma^2}} + \frac{\varepsilon}{\Delta}}.$$

Таким образом после деления на числитель и нормировки на единицу в нуле мы получим выражение для оптимальных весов

$$w_{\text{opt}}(e) = \frac{1+c}{1+c \exp\left(\frac{e^2}{2\sigma^2}\right)}, \quad c_{\text{opt}} = \frac{\varepsilon\sigma\sqrt{2\pi}}{\Delta(1-\varepsilon)}$$

с константой

Поскольку $\sigma \ll \Delta$, эта константа довольно мала даже при малом отношении сигнал/шум.

Таким образом мы получили уравнение правдоподобия (3.7), аналогичное нормальному уравнению МНК (3.6), но с заменой числовых весовых коэффициентов на **весовые функции $w(e_i)$** , которые приходится перевычислять на каждой итерации получившейся итеративной процедуры, названной процедурой Флетчера-Гранта-Хеблена (ФГХ). На каждой итерации ФГХ-процедуры выполняется взвешенный МНК, но с функциональными весами. Если нет каких-либо априорных соображений по выбору начальных значений весов, то можно инициировать ФГХ процедуру с помощью обычного МНК, взяв единицы в качестве весов на нулевой итерации: $w_i^{(0)} \equiv 1$.

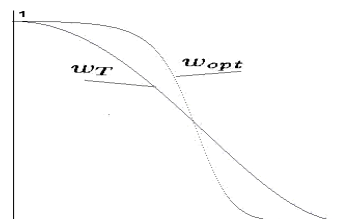
Из соображений ускорения вычислений можно воспользоваться полиномиальной аппроксимацией функции $w_{opt}(\mathbf{e})$ многочленом 4-го порядка

$$w(e) = \begin{cases} \left(1 - \left(\frac{e}{c_T \sigma}\right)^2\right)^2, & \text{если } |e| < c_T \sigma, \\ 0 & \text{в остальных случаях.} \end{cases}$$

который известен также как бивес Тьюки.

Оптимальная весовая функция и бивес Тьюки изображены на рисунке.

Бивесовая функция используется весьма часто, хотя следует



подчеркнуть необходимость тщательного выбора параметра c_T . Дж.Тьюки рекомендовал $c_T = 3 \div 5$. При этом следует учитывать, что выброс измерений с большими отклонениями путем приписывания им нулевых весов идет по величине σc_T , где значение σ также подлежит перевычислению на каждой k -й итерации по формуле

$$\sigma^{(k)} = \sqrt{\frac{\sum_i w_i^{(k-1)} (e_i^{(k-1)})^2}{\sum_i w_i^{(k-1)}}}.$$

Следует также помнить, что применение робастного оценивания параметров имеет смысл только в случаях засоренности выборки. Если, напротив, есть уверенность, что выборка свободна от фоновых измерений, то никакой необходимости в использовании итеративной процедуры с функциональными весами нет, т.к. для таких случаев доказана максимальная эффективность обычного МНК.

При решении задачи в пакетах EXCEL - STATISTICA применяется более просто вычисляемая весовая функция типа распределения Коши:

$w(d) = 1/(1+\alpha*d^2)$ с малым коэффициентом $\alpha=0.1—0.5$, подбираемом в зависимости от числа точек-«выбросов»

В программе STATISTICA в окне «Nonlinear Estimation» («Нелинейное Оценивание») выбрать пункт «User-specified Regression, custom loss function» («Регрессия с Пользовательской Функцией Потерь»);

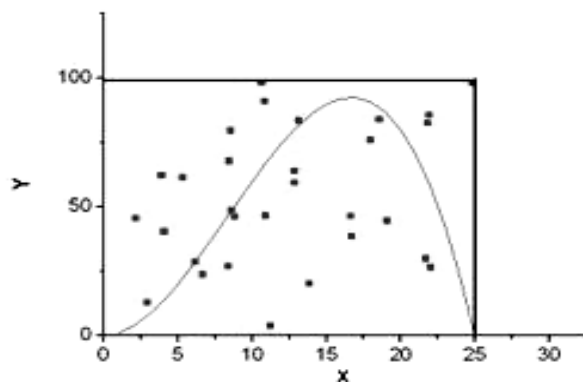
• после нажатия кнопки «Function to be estimated & loss function» («Оцениваемая Функция и Функция Потерь») указываем в первом поле ввода формулу $v_2=a*v_1+b$, а во втором – выражение $(OBS-PRED)**2/(1 + 0.2*(OBS-PRED)**2)$, которая соответствует формуле $w(d) = 1/(1+\alpha*d^2)$ при $\alpha=0.2$

Тема 4. Понятие об интегрировании методом Монте-Карло.

Чтобы взять интеграл от функции $f(x)$ на отрезке (a,b) ,

$$I = \int_a^b f(x)dx$$

воспользуемся неформальным геометрическим описанием интеграла и будем понимать его как площадь под графиком этой функции. Для определения этой площади можно использовать следующий стохастический алгоритм:



- ограничим функцию прямоугольником (n -мерным параллелепипедом в случае многих измерений), площадь которого S_{par} легко вычисляется как $(b-a)M$, где $M = \max_{(a,b)} f(x)$;

- «набросаем» в этот прямоугольник (параллелепипед) некоторое количество точек (n штук), координаты которых будем выбирать случайным образом;
- определим число точек (m штук), которые попадут под график функции;
- площадь функции S , равная искомому интегралу I , будет приблизительно равна величине

$$I_n = S_{par} \frac{m}{n} .$$

Для определения точности, с которой был вычислен интеграл I методом Монте-Карло, т.е. величины $|I_n - I|$ следует учесть, что оценка I_n является случайной величиной. Поэтому для некоторой малой величины $\varepsilon > 0$ мы можем оценить только вероятность события, что наша оценка окажется слишком грубой $|I_n - I| > \varepsilon$, ограничив эту вероятность другой малой величиной α : $P\{|I_n - I| > \varepsilon\} < \alpha$.

Событие (успех), состоящее в попадание точки под кривую, происходит с вероятностью $p = I / S_{par}$. Опишем очередное из этих событий вспомогательной случайной величиной ξ_i , которая с вероятностью p равна 1, и нулю в противном случае. Нетрудно подсчитать математическое ожидание и дисперсию ξ_i : $M\xi_i = p$, $D\xi_i = p - p^2 = p(1 - p) = pq$. Очевидно, что число таких событий (число m успехов в серии из n испытаний) будет иметь биномиальное распределение со средним np и дисперсией npq . Отсюда мы можем найти математическое ожидание и дисперсию интересующей нас оценки I_n интеграла I : $M(I_n) = M(S_{par} m/n) = S_{par} M(m)/n = p S_{par} = I$

(что говорит о несмещенности оценки);

$D(I_n) = D(S_{par} m/n) = (S_{par}/n)^2 npq = S_{par}^2 pq/n$, а также и

среднеквадратичную величину $\sigma I_n = S_{par} \sqrt{\frac{pq}{n}}$.

Выберем теперь для оценки очень малой вероятности величину $\alpha = 1\%$, Тогда вероятность того, что отклонение интеграла I от его оценки I_n не превысит ε будет равна $P\{|I_n - I| < \varepsilon\} = 1 - \alpha = 0.99$.

Далее мы учтем, что в соответствии с центральной

предельной теоремой теории вероятностей при

больших n величина I_n как сумма большого числа

слагаемых имеет приблизительно нормальное

распределение, так что после ее нормировки мы

получаем: $P\{|I_n - I| < \varepsilon\} = P\left\{\frac{|I_n - I|}{S_{par} \sqrt{\frac{pq}{n}}} < \frac{\varepsilon}{S_{par} \sqrt{\frac{pq}{n}}}\right\} = \frac{2}{\sqrt{2\pi}} \int_0^\beta e^{-\frac{x^2}{2}} dx,$

где обозначено $\beta = \frac{\varepsilon}{S_{par} \sqrt{\frac{pq}{n}}}$.

Чтобы последний интеграл стал равен $1 - \alpha = 0.99$, мы должны выбрать $\beta = 2.6$, откуда получаем искомую оценку точности нашего интегрирования в виде

$$\varepsilon = \beta S_{par} \sqrt{\frac{pq}{n}}. \quad (4.1)$$

Величины $p=I/S_{par}$ и $q=1-p$ нам неизвестны, но мы можем оценить p , используя вычисленную оценку интеграла I_n , как. $\hat{p} = I_n / S_{par}$.

Например, при вычислении по 100 точкам известного интеграла $I = \int_0^1 x^2 dx = \frac{1}{3}$

мы получаем число $I_n = 0.32$. В этом случае $S_{par} = 1$ и мы получаем довольно грубую оценку точности как $\varepsilon = 2.6 \sqrt{\frac{0.32 \cdot 0.68}{100}} \cong 0.12$.

Эту же формулу (4.1) можно использовать для приблизительной оценки числа испытаний, необходимого для достижения заданной точности.

Выражая n из (4.1), получаем $n \geq pq \left(\frac{\beta S_{par}}{\varepsilon} \right)^2$, что при $\varepsilon = 0.01$ в условиях предыдущего примера приводит к требованию - $n \sim 14710$.

Формула (4.1) с ее оценкой точности вида $\frac{const}{\sqrt{n}}$ при небольших размерностях области интегрирования заметно уступает обычным кубатурным методам интегрирования. Однако начиная с размерностей 3 – 4 и выше, монте-карловское интегрирование оказывается более точным, особенно в случаях, когда функция задается неявно, а область интегрирования представляется в виде сложных неравенств.