Про погрешности аппроксимации

Ахундзянов Амир Андреевич

12 ноября 2021 г.

1. Введение

2. Общие факты

2.1. Определение погрешности

Под погрешностью величины x будет подразумеваться среднеквадратичное отклонение

$$\sigma_x = M[(x - M[x])^2] = M[\Delta x^2] = \langle (x - \langle x \rangle)^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2$$

где $M[x]$ - матожидание, $\langle x \rangle$ - среднее. То есть матожидание квадрата отклонения от среднего или средний квадрат отклонения в случае конечной равновероятной выборки. Последняя формула получается просто раскрытием скобок.

Если распределение x нормальное, среднеквадратичное отклонение определяет коэффициент в показателе экспоненты

$$P(x) \propto \exp\left(\frac{-\Delta x^2}{2\sigma_x^2}\right)$$

2.2. Формула тейлора

Функцию одной переменной можно приблизить многочленом Тейлора

$$f(x_0 + \Delta x) = f(x_0) + f^{(1)}(x_0)\Delta x + ... + \frac{f^{(n)}(x_0)\Delta x^n}{n!} + o(\Delta x^n)$$

Функцию нескольких переменных можно разложить по одной и переменных

$$f(x_0 + \Delta x, y) = f(x_0, y) + \frac{\partial f(x_0, y)}{\partial x}\Delta x + ... + \frac{\partial^n f(x_0, y)}{\partial x^n}\frac{\Delta x^n}{n!} + o(\Delta x^n)$$

А потом каждое слогаемое разложить в ряд Тейлора по второй переменной. То есть считая Δx и Δy одного порядка малости (все частные производные в точке (x_0, y_0) и используется обозначение $\partial_x f = \frac{\partial f}{\partial x}$)

$$f(x_0 + \Delta x, y_0 + \Delta y) =$$

$$f + \quad \partial_x f \Delta x \quad + ...+ \partial_x^i f \frac{\Delta x^i}{i!} \quad + ... + \partial_x^n f \frac{\Delta x^n}{n!} + o(\Delta x^n) +$$

$$+\partial_y f \Delta y + \partial_y \partial_x f \Delta x \Delta y + ...+ \partial_y \partial_x^i f \frac{\Delta x^i \Delta y}{i!} +$$

$$...$$

$$+...+\partial_y^j \partial_x^i f \frac{\Delta x^i \Delta y^j}{i!j!} + ...+$$

$$...$$

$$+\partial_y^n f \frac{\Delta y^n}{n!}$$

То же самое можно переписать, как суммы слагаемых с одинаковой суммой степеней при Δи $x\Delta y$

$$f(x_0 + \Delta x, y_0 + \Delta y) = \sum_{k=0}^{n} \sum_{i=0}^{k} \partial_x^i \partial_y^{k-i} f \frac{\Delta x^i \Delta y^{k-i}}{i!(k-i)!} + o(\Delta x^n)$$

Можно еще вынести $k!$ за скобку и получатся коэффициенты бинома ньютона

$$f(x_0 + \Delta x, y_0 + \Delta y) = \sum_{k=0}^{n} \frac{1}{k!} \sum_{i=0}^{k} \partial_x^i \partial_y^{k-i} f \Delta x^i \Delta y^{k-i} C_k^i + o(\Delta x^n)$$

Нас в частности будет интересовать разложение до второго порядка малости

$$f(x_0 + \Delta x, y_0 + \Delta y) = f(x_0, y_0) + \partial_x f(x_0, y_0) \Delta x + \partial_y f(x_0, y_0) \Delta y +$$
$$+ \frac{1}{2} \partial_x^2 f(x_0, y_0) \Delta x^2 + \partial_x \partial_y f(x_0, y_0) \Delta x \Delta y + \frac{1}{2} \partial_y^2 f(x_0, y_0) \Delta y^2$$

2.3. Метод максимального правдоподобия

4.2. Метод максимального правдоподобия.

Рассмотрим кратко один из наиболее общих методов оценки параметров зависимостей — метод максимума правдоподобия.

Сделаем два ключевых предположения:

- зависимость между измеряемыми величинами действительно может быть описана функцией $y = f(x \,|\, \theta)$ при некотором θ;
- все отклонения Δy_i результатов измерений от теоретической модели являются *независимыми* и имеют *случайный* (не систематический!) характер.

Пусть $P(\Delta y_i)$ — вероятность обнаружить отклонение Δy_i при фиксированных $\{x_i\}$, погрешностях $\{\sigma_i\}$ и параметрах модели θ. Построим функцию, равную вероятности обнаружить весь набор отклонений $\{\Delta y_1, \ldots, \Delta y_n\}$. Ввиду независимости измерений она равна произведению вероятностей:

$$L = \prod_{i=1}^{n} P(\Delta y_i). \tag{4.2}$$

Функцию L называют *функцией правдоподобия.*

Метод максимума правдоподобия заключается в поиске такого θ, при котором наблюдаемое отклонение от модели будет иметь *наибольшую вероятность*, то есть

$$L(\theta) \to \max.$$

Замечание. Поскольку с суммой работать удобнее, чем с произведениями, чаще используют не саму функцию L, а её логарифм:

$$\ln L = \sum_i \ln P(\Delta y_i).$$

Пусть теперь ошибки измерений имеют *нормальное* распределение. Согласно (2.5), вероятность обнаружить в i-м измерении отклонение Δy_i пропорциональна величине

$$P(\Delta y_i) \propto e^{-\frac{\Delta y_i^2}{2\sigma_i^2}},$$

где σ_i — стандартная ошибка измерения величины y_i. Тогда логарифм функции правдоподобия (4.2) будет равен (с точностью до константы)

$$\ln L = -\sum_i \frac{\Delta y_i^2}{2\sigma_i^2} = -\frac{1}{2}\chi^2.$$

Таким образом, максимум правдоподобия действительно будет соответствовать минимуму χ^2.

Фрагмент отсюда П.В. Попов, А.А. Нозик *Обработка результатов учебного эксперимента*

2.4. Погрешность функции

2.6.2. Случай многих переменных

Пусть величина u вычисляется по измеренным значениям нескольких различных *независимых* физических величин x, y, ... на основе известного закона $u = f(x,y,\ldots)$. В качестве наилучшего значения можно по-прежнему взять значение функции f при наилучших значениях измеряемых параметров:

$$u^\star = f(x^\star,y^\star,\ldots)\,.$$

Для нахождения погрешности σ_u воспользуемся свойством, известным из математического анализа, — малые приращения гладких функции многих переменных складываются линейно, то есть справедлив *принцип суперпозиции* малых приращений:

$$\Delta u \approx f'_x \cdot \Delta x + f'_y \cdot \Delta y + \ldots,$$

где символом $f'_x \equiv \frac{\partial f}{\partial x}$ обозначена *частная производная* функции f по переменной x — то есть обычная производная f по x, взятая при условии, что все остальные аргументы (кроме x) считаются постоянными параметрами. Тогда пользуясь формулой для нахождения дисперсии суммы независимых величин (2.7), получим соотношение, позволяющее вычислять погрешности косвенных измерений для произвольной функции $u = f\,(x,y,\ldots)$:

$$\boxed{\sigma_u^2 = f_x'^2\,\sigma_x^2 + f_y'^2\,\sigma_y^2 + \ldots} \tag{2.11}$$

Отметим, что формулы (2.10) и (2.11) применимы только если относительные отклонения всех величин малы ($\varepsilon_x,\varepsilon_y,\ldots \ll 1$), а измерения проводятся вдали от особых точек функции f (производные f'_x, f'_y ... не должны обращаться в бесконечность). Также подчеркнём, что все полученные здесь формулы справедливы только для *независимых* переменных x, y, ...

Взято оттуда же. Убедиться в том, на сколько этот вывод некорректный можно рассмотрев такой пример.

$$f(x,y) = \frac{x}{y}$$

Имея измерение $t \pm \Delta t$, рассчитаем значение

$$f(t,t) \equiv 1$$

Понятно, что погрешность этого результата нулевая, а формула даст нам ненулевую. Далее будет показано, как в случае зависимых переменных считать погрешность функции, а при определении параметров методом аппроксимации они получаются зависимыми. Также будет показано, как считаются погрешности параметров и их дисперсии.

3. Основная часть

3.1. Дисперисия параметров аппроксимации

4.5. Оценка погрешности параметров

Важным свойством метода хи-квадрат является «встроенная» возможность нахождения погрешности вычисленных параметров σ_θ.

Пусть функция $L(\theta)$ имеет максимум при $\theta = \hat{\theta}$, то есть $\hat{\theta}$ — решение задачи о максимуме правдоподобия. Согласно центральной предельной теореме мы ожидаем, что функция правдоподобия будем близка к нормальному распределению: $L(\theta) \propto \exp\left(-\frac{(\theta-\hat{\theta})^2}{2\sigma_\theta^2}\right)$, где σ_θ — искомая погрешность параметра. Тогда в окрестности $\hat{\theta}$ функция $\chi^2(\theta) = -2\ln(L(\theta))$ имеет вид параболы:

$$\chi^2(\theta) = \frac{(\theta-\hat{\theta})^2}{\sigma_\theta^2} + \mathrm{const}.$$

Легко убедиться, что:

$$\chi^2(\hat{\theta} \pm \sigma_\theta) - \chi^2(\hat{\theta}) = 1.$$

Иными словами, при отклонении параметра θ на одну ошибку σ_θ от значения $\hat{\theta}$, минимизирующего χ^2, функция $\chi^2(\theta)$ изменится на единицу. Таким образом для нахождения *интервальной* оценки для искомого параметра достаточно графическим или численным образом решить уравнение

$$\Delta\chi^2(\theta) = 1. \tag{4.4}$$

Вероятностное содержание этого интервала будет равно 68% (его еще называют 1–σ интервалом). Отклонение χ^2 на 2 будет соответствовать уже 95% доверительному интервалу.

Отмечу также, что среднеквадратичное отклонение можно получить взяв вторую производную

$$\frac{\partial^2\chi^2(\theta)}{\partial\theta^2} = \frac{2}{\sigma_\theta^2}$$

Теперь рассмотрим случай нескольких переменных. По определению запишем погрешность функции от нескольких переменных.

$$\begin{aligned}\sigma_f^2 = M[(f(x+\Delta x, y+\Delta y) - f(x,y))^2] &\approx M[(\partial_x f\Delta x + \partial_y f\Delta y)^2] \\ = M[(\partial_x f)^2\Delta x^2] &+ M[2\partial_x f\Delta x\partial_y f\Delta y] + M[(\partial_y f)^2\Delta y^2] \\ = (\partial_x f)^2 M[\Delta x^2] &+ 2\partial_x f\partial_y f M[\Delta x\Delta y] + (\partial_y f)^2 M[\Delta y^2] \\ = (\partial_x f)^2\sigma_x^2 &+ 2\partial_x f\partial_y f M[\Delta x\Delta y] + (\partial_y f)^2\sigma_y^2\end{aligned}$$

Остаётся множитель $M[\Delta x\Delta y]$, который не выражается через дисперсии x и y. Он показывает связь между природой погрешности x и y. Называется это ковариацией. В случае, если ошибки независимые,

$$M[\Delta x\Delta y] = M[\Delta x]M[\Delta y] = 0\cdot 0$$

Разберёмся как можно рассчитать ковариацию параметров аппроксимации.

3.2. Рассчёт ковариации

Разложим $\chi^2(\theta_i)$ функцию всех определяемых параметров вблизи минимума как функцию от двух параметров (θ_1 и θ_2) по формуле Тейлор до первого неисчезающего порядка. Будем считать, что значение параметров соответствующее минимуму равно нулю, чтобы не писать лишний раз Δ Как было показано выше разложение будет иметь такой вид

$$\chi^2(x,y) \approx const + \frac{\partial^2\chi^2(x,y)}{\partial x^2}\frac{\Delta x^2}{2} + \frac{\partial^2\chi^2(x,y)}{\partial y^2}\frac{\Delta y^2}{2} + \frac{\partial\chi^2(x,y)}{\partial x}\frac{\partial\chi^2(x,y)}{\partial y}\Delta x\Delta y$$

Подставим выражение для второй производной через σ и запишем множитель перекрёсного слагаемого через обозначение

$$corr = \frac{\frac{\partial\chi^2}{\partial x}\frac{\partial\chi^2}{\partial y}}{\sqrt{\frac{\partial^2\chi^2}{\partial x^2}\frac{\partial^2\chi^2}{\partial y^2}}} = -\frac{\partial_{\theta_1}\chi^2\partial_{\theta_2}\chi^2}{\sqrt{\partial^2_{\theta_1}\chi^2\partial^2_{\theta_2}\chi^2}} = \frac{D_{\theta_1\theta_2}}{\sigma_{\theta_1}\sigma_{\theta_2}}$$

($D_{\theta_1\theta_2} = -\frac{\partial_{\theta_1}\chi^2\partial_{\theta_2}\chi 2}{2}$ тоже просто новое обозначение)

$$\chi^2 - const = \frac{\theta_1^2}{\sigma_{\theta_1}^2} - 2\frac{corr\theta_1\theta_2}{\sigma_{\theta_1}\sigma_{\theta_2}} + \frac{\theta_2^2}{\sigma_{\theta_2}^2}$$

Очень хочется сказать, что это квадрат суммы, но там перекрёстное слагаемое другое, так что естественно будет сказать, что это скалярный квадрат суммы двух векторов. Тогда угол между ними будет определять множитель этого перекрёстного слагаемого. Чтобы было совсем красиво, возьмём эти базисные вектора не только не ортогональными, но и не нормированными. Тогда

$$|\vec{e}_{\theta_1}| = \frac{1}{\sigma_{\theta_1}}$$

$$\vec{\theta_1} = \theta_1 \vec{e}_{\theta_1}$$

$$|\vec{e}_{\theta_2}| = \frac{1}{\sigma_{\theta_2}}$$

$$\vec{\theta_2} = \theta_2 \vec{e}_{\theta_2}$$

$$\frac{corr}{\sigma_{\theta_1}\sigma_{\theta_2}} = -\vec{e}_{\theta_1}\vec{e}_{\theta_2} = \frac{-\cos(\alpha)}{\sigma_{\theta_1}\sigma_{\theta_2}}$$

$$\frac{\theta_1^2}{\sigma_{\theta_1}^2} - 2\frac{corr\theta_1\theta_2}{\sigma_{\theta_1}\sigma_{\theta_2}} + \frac{\theta_2^2}{\sigma_{\theta_2}^2} = \vec{\theta_1}^2 + 2\vec{\theta_1}\vec{\theta_2} + \vec{\theta_2}^2 = (\vec{\theta_1} + \vec{\theta_2})^2$$

Тогда в ортонормированном базисе(x, y), где косинус угла между векторами $\vec{e}_{\theta_1}$ и $\vec{e}_{\theta_2}$ равен $corr$ выражение для χ^2 равно просто скалярному квадрату вектора к точке, что есть квадрату расстояния до начала координат. Тогда распределение вероятностей в этих координатах равное $L = exp(-\frac{\chi^2}{2})$ сферически симметричное. Тогда можно посчитать среднее на окружностях в таком пространстве и усреднить это по всем окружностям. Найдём значения θ_1 и θ_2 как функцию новых координат.

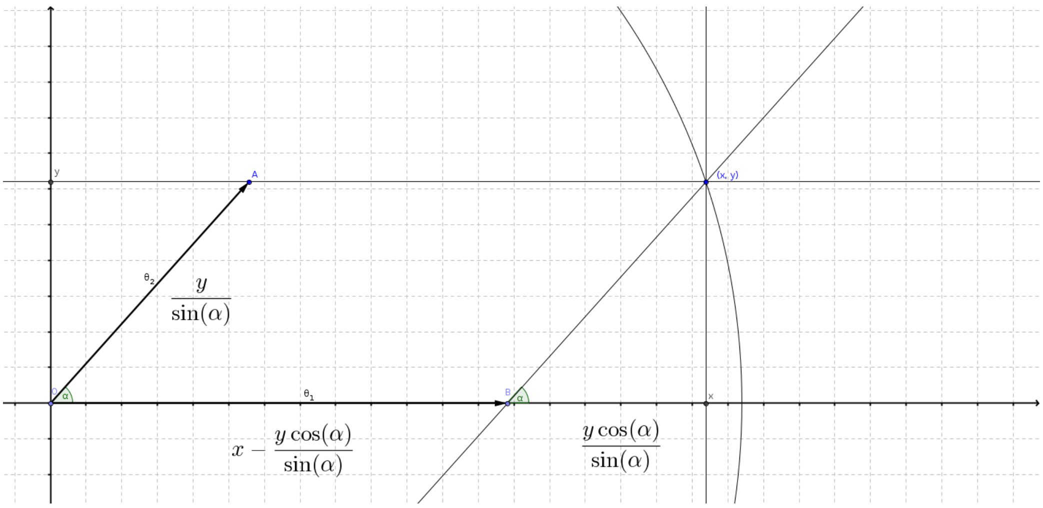

Рис. 1. Рисунок в ортонормированном базисе (x, y)

Возьмем точку (x, y) и разложим ее по базису $\vec{e}_{\theta_1}$, $\vec{e}_{\theta_2}$ (см. рис. 1).

Из простых геометрических соображений выражаем длины векторов $\vec{\theta_1}$ и $\vec{\theta_2}$ через координаты и α - угол между единичными векторами.

$$\left|\vec{\theta_2}\right| = |\vec{e}_{\theta_2}|\,\theta_2 = \frac{y}{\sin(\alpha)}$$

$$\left|\vec{\theta_1}\right| = |\vec{e}_{\theta_1}|\,\theta_1 = x - \frac{y\cos(\alpha)}{\sin(\alpha)} = x - \cos(\alpha)\left|\vec{\theta_2}\right|$$

Теперь нужно усреднить по окружности выражение

$$\langle\theta_1\theta_2\rangle = \sigma_1\sigma_2\langle\left|\vec{\theta_2}\right|\left|\vec{\theta_1}\right|\rangle = \sigma_1\sigma_2\langle\left|\vec{\theta_2}\right|\left(x - \cos(\alpha)\frac{\theta_2}{\sigma_2}\right)\rangle = \sigma_1\sigma_2\left(\frac{\langle xy\rangle}{\sin\alpha} - \cos(\alpha)\frac{\langle\theta_2^2\rangle}{\sigma_2^2}\right)$$

Теперь нужно заметить, что среднее $\langle xy\rangle$ равно нулю. Это понятно например из того, что при каждом x распределение вероятности y симметрично относительно нуля. Среднее от произведения $\langle\theta_2^2\rangle$ при усреднении сначала по окружности, а потом по всем таким окружностям будет тем же, чем и при усреднении по всему вероятностному пространству, то есть просто среденеквадратичным отклонением (мы договорились, что у θ_2 отклонение от среднего равно значению). Тогда

$$\langle\theta_1\theta_2\rangle = -\cos(\alpha)\sigma_1\sigma_2\frac{\sigma_2^2}{\sigma_2^2} = \sigma_1\sigma_2 corr = D_{\theta_1\theta_2}$$

В последних равенствах просто подставлены обозначения введенные выше по определению. Таким образом мы выразили ковариацию двух параметров через коэффициенты разложения χ^2 вблизи минимума

Далее $D_{xy} = \langle\Delta x\Delta y\rangle = \langle(x - \langle x\rangle)(y - \langle y\rangle)\rangle = \langle xy\rangle - \langle x\rangle\langle y\rangle$ будет использоваться как обозначение ковариации x и y, а дисперсию будем обозначать как ковариацию величины с самой собой $D_{xx} = \sigma_x^2$

Перепишем ковариацию через начальные частные производные

$$D_{\theta_1\theta_2} = -\frac{\partial_{\theta_1}\chi^2\partial_{\theta_2}\chi^2}{2}$$

То есть умея брать производную у χ^2 сможем найти и ковариацию параметров аппроксимации. Можно также по аналогии с методом нахождения среднеквадратичного отклонению по приравниванию увеличения χ^2 относительно минимума к 1 приравнять изменение $\chi^2(\theta_1, \theta_2))$ относительно минимума к 1 получить численное решение уравнения

$$1 = \frac{\theta_1^2}{D_{11}} - 2\frac{D_{12}\theta_1\theta_2}{D_{11}D_{22}} + \frac{\theta_2^2}{D_{22}}$$

Где символ θ в индексах опущен для большей наглядности. Решением будет эллипс, по параметрам которого при большом желании тоже можно будет оценить параметры и получить ковариацию.

4. Погрешность функции от зависимых переменных

В случае зависимх переменных слогаемые с ковариациями ($\langle\theta_1\theta_2\rangle$) не будет нулевым. Распишем по определению погрешность через линейное приближение функции.

$$\langle\Delta f(\theta)^2\rangle = \langle(f(\theta+\Delta\theta)-f(\theta))^2\rangle = \langle(\sum_{i=1}^{n}\Delta f(\Delta\theta_i))^2\rangle = \langle\sum_{i=1}^{n}\sum_{j=1}^{n}\Delta f(\Delta\theta_i)\Delta f(\Delta\theta_j)\rangle =$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n}\langle\Delta\theta_i\Delta\theta_j\rangle\partial_{\theta_i}f\partial_{\theta_j}f = \sum_{i=1}^{n}\sum_{j=1}^{n}D_{\theta_i\theta_j}\partial_{\theta_i}f\partial_{\theta_j}f$$

Где сначала $\Delta\theta_i$ множество параметров, где i-ый элемент изменён, а остальные нет, а потом просто изменение i-ого аргумента, $\Delta f(\Delta\theta_i) = f(\Delta\theta_i) - f(\theta_i)$.

При обработке данных может быть удобнее использовать формулу через численную частную производную и корреляцию.

$$\sum_{i=1}^{n}\sum_{j=1}^{n}corr_{ij}\Delta f(\Delta\theta_i)\Delta f(\Delta\theta_j)$$

Где уже $\Delta\theta_i$ множество параментов θ в котором i-ый параметр изменили на σ_{θ_i}

5. Вывод