

Компьютерная обработка результатов измерений

Лекция 3. Теория физических измерений. Систематические и случайные погрешности.
Лекция 4. Теория оценок.

Емельянов Эдуард Владимирович

Специальная астрофизическая обсерватория РАН
Лаборатория обеспечения наблюдений

22 марта 2021 года



- 1 Измерения и величины
- 2 Погрешность
- 3 Метод наименьших квадратов
- 4 Правило «трех сигм»
- 5 Иррегулярно распределенные данные



Измерения и величины

Мерой называется средство измерений, предназначенное для воспроизведения и хранения значения физической величины. Результатом сравнения оцениваемой вещи с мерой является именованное число, называемое **значением величины**.

Физические величины

- постоянные (инварианты, константы, априорно фиксированные значения);
- изменяющиеся (по определенному закону от t);
- случайные (не имеющие точного значения).

Единицы измерения, размерность.

Скалярные, векторные, комплексные, тензорные величины.

Метрология.



Виды измерений

Прямое при котором искомое значение физической величины получают непосредственно.

Косвенное на основании результатов прямых измерений других физических величин, функционально связанных с искомой величиной.

Совместные проводимые одновременно для нескольких неодноименных величин для определения зависимости между ними.

Совокупные при которых искомые значения величин определяют путем решения системы уравнений, получаемых при измерениях этих величин в различных сочетаниях.

Равноточные выполненные одинаковыми по точности средствами измерений.

Неравноточные выполненных различающимися по точности средствами измерений и (или) в разных условиях.



Виды измерений

Однократное, многократное

Статическое для величины, принимаемой в соответствии с конкретной измерительной задачей за неизменную на протяжении времени измерения.

Динамическое для изменяющейся по размеру физической величины.

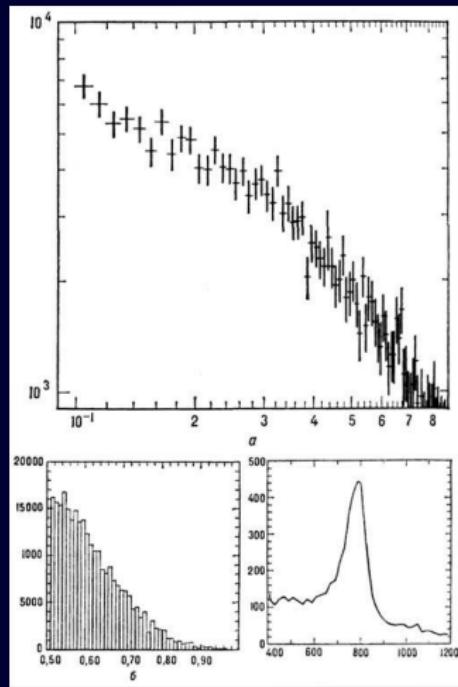
Абсолютное основанное на прямых измерениях одной или нескольких основных величин и (или) использовании значений физических констант.

Относительное сравнение с эталонными мерами.



Представление результатов измерений

Графическое



Табличное

	A	B	C	D	E
1	Доходы от ярмарки				
2	Товар	Количество	Цена	Общая стоимость	Доля от прибыли в %
3	Сдобные булочки	130	10	1300	13%
4	Пирожки	130	10	1300	13%
5	Соки на разлив	140	10	1400	14%
6	Разделочные доски	50	50	2500	25%
7	Швабры	15	100	1500	15%
8	Табуретки	10	200	2000	20%
9	Итого:			10000	



Погрешность

Погрешность — отклонение измеренного значения величины от её истинного (действительного) значения.

Абсолютная погрешность, Δx (напр., RMS); относительная погрешность, $\delta x = \Delta x / \bar{x}$; приведенная погрешность $\gamma x = \Delta x / N_x$ (нормировочный коэффициент).

По причине возникновения

Инструментальные определяются погрешностями применяемых средств измерений.

Методические обусловлены несовершенством метода, а также упрощениями,ложенными в основу методики.

Субъективные обусловлены качествами экспериментатора.



Погрешность

По характеру проявления

Случайные обусловлены совокупностью внешних факторов, влияющих на результат эксперимента.

Систематические связаны с влиянием прибора на измеряемую величину или методическими ошибками, выявляются лишь сменой прибора/метода/экспериментатора.

Промахи наиболее сильно себя проявляют и связаны с неисправностью прибора или экспериментатора.

Средняя квадратическая погрешность среднего арифметического

$$\sigma_{\langle x \rangle} = \frac{\sigma_x}{\sqrt{n}} = \sqrt{\frac{\sum_{i=1}^n (\bar{x}_i - \langle x \rangle)^2}{n(n-1)}}.$$



Доверительный интервал

Доверительная вероятность

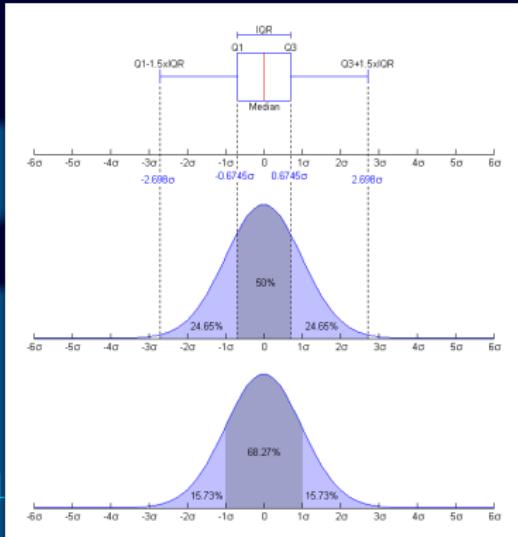
$$p = P(X_0 \leq x \leq X_1)$$

Математическое ожидание

Если известен закон распределения (мат. ожидание и дисперсия: μ и σ), то

$$P\left(\langle X \rangle - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \langle X \rangle + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1$$

где z_α – α -квантиль нормального распределения



Квартили: первый (0.25-квантиль), второй (0.5-квантиль, медиана) и третий (0.75-квантиль).



Доверительный интервал

α -квантилем называется число x_α : $P(X \leq x_\alpha) \geq \alpha$ и $P(X \geq x_\alpha) \geq 1 - \alpha$.
Т.е. по интегральной функции распределения $F(x_\alpha) = \alpha$. А т.к.
 $P(a \leq X \leq b) = F(b) - F(a)$, получаем:

$$P(x_{1-\frac{\alpha}{2}} \leq X \leq x_{1+\frac{\alpha}{2}}) = \alpha.$$

Пример

В 64 наблюдениях получено: $S_1 = \sum x = 600$, $S_2 = \sum(x - \bar{x})^2 = 3800$.

Вычислить 90% доверительный интервал матожидания.

Решение: $\sigma = \sqrt{S_2/(n - 1)} = 7.72$; $\langle x \rangle = S_1/n = 9.375$. $F(0.05) = 1.96$,
отсюда найдем границы интервала $\langle x \rangle \pm F(0.05)\sigma/\sqrt{n}$: $\bar{x} \in [7.484, 11.266]$ с
точностью 90%.



Доверительный интервал

Математическое ожидание

Если закон распределения неизвестен, то

$$P\left(\bar{X} - t_{1-\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{1-\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}}\right) = 1 - \alpha,$$

где S – несмешенный RMS. Величина

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

имеет распределение Стьюдента, а $t_{\alpha, n-1}$ – его квантили.

Пример: $\bar{X} = 10$, $S_n = 2$, $n = 11$ (10 степеней свободы), по таблице для двухстороннего распределения Стьюдента с вероятностью 95% $T_{10}^{95} = 2.228$. Тогда доверительный интервал есть $\bar{X} \pm TS_n/\sqrt{n}$, т.е. $\mu \in (8.6565, 11.3440)$.



Доверительный интервал

Дисперсия

Если известно среднее, можно воспользоваться распределением χ^2 .

$$P\left(\frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{(1+\alpha)/2,n}^2} \leqslant \sigma^2 \leqslant \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{(1-\alpha)/2,n}^2}\right) = \alpha.$$

Если же среднее неизвестно, то

$$P\left(\frac{(n-1)S^2}{\chi_{(1+\alpha)/2,n-1}^2} \leqslant \sigma^2 \leqslant \frac{(n-1)S^2}{\chi_{(1-\alpha)/2,n-1}^2}\right) = \alpha.$$



Правила вычисления погрешностей

1

$$\Delta(\sum a_n) = \sum \Delta a_n.$$

2

$$\prod(a_i \pm \Delta a_i) = \prod a_i \prod(1 \pm \delta a_i) \approx \prod a_i (1 \pm \sum \delta a_i),$$
$$(a[1 \pm \delta a])^n \approx a^n (1 \pm n\delta a).$$

3 В сложных функциях вида $y = f(x_1, \dots, x_n)$ можно оценить погрешность, воспользовавшись приближением:

$$\delta y \approx \left| \frac{dy}{y} \right| = \left| \frac{df(x_1, \dots, x_n)}{f(x_1, \dots, x_n)} \right|,$$

в котором следует заменить $dx_i/x_i = \delta x_i$ – относительная погрешность измерения величины x_i , $dx_i = \Delta x_i$ – абсолютная погрешность. Все слагаемые необходимо суммировать по абсолютной величине.

Метод наименьших квадратов

Пусть имеется функция $f(x|a)$, зависящая от аргумента x и набора параметров a . Данной функции соответствует набор пар данных (x_n, y_n) , причем $y_n = f(x_n|a) + \varepsilon_n$, где ε_n – случайная ошибка. Математическое ожидание ошибки $\bar{\varepsilon} = 0$, ее среднеквадратическое отклонение равно σ_n . Для оценки a (аппроксимации набора данных заданной функцией) необходимо минимизировать выражение

$$\Phi = \sum_{n=1}^N \left(\frac{y_n - f(x_n|a)}{\sigma_n^2} \right)^2.$$

При этом подразумевается, что число измерений превышает число параметров a .



Пример: линейная зависимость

Пусть $y = ax + b$, x_i известны с пренебрежимо малой погрешностью, y_i – результаты измерений с нормальным распределением, $\bar{y}_i = ax_i + b$.

Минимизируем величину $Y = \sum(y_i - \bar{y}_i)^2$, $\frac{\partial Y}{\partial a} = 0$, $\frac{\partial Y}{\partial b} = 0$:

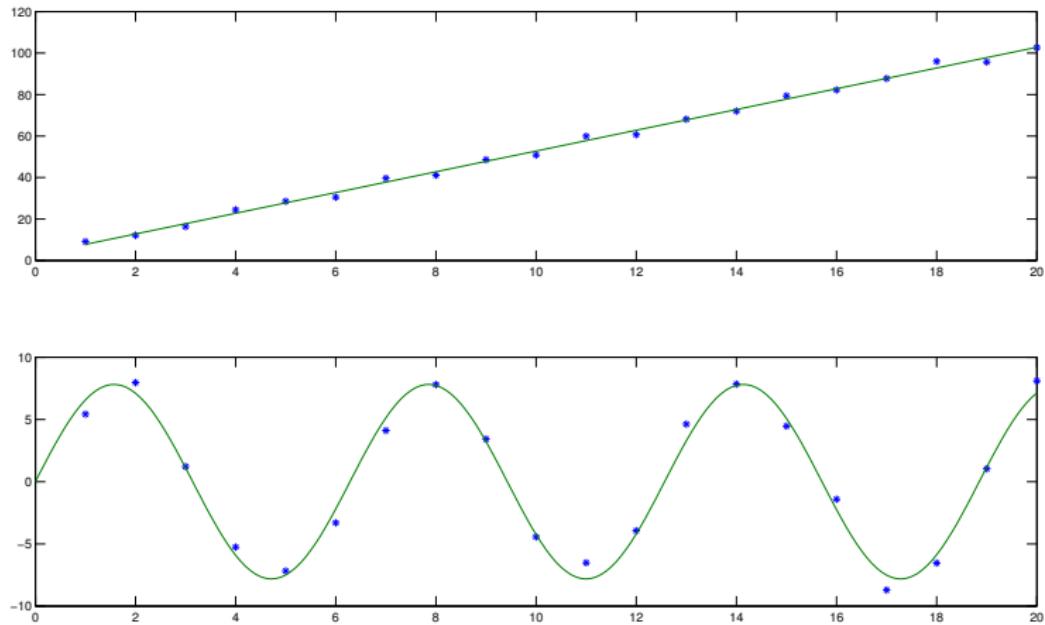
$$a = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} = \frac{\bar{x}\bar{y} - \bar{x}\bar{y}}{\bar{x}^2 - (\bar{x})^2},$$

$$b = \frac{\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i}{n \sum x_i^2 - (\sum x_i)^2} = \frac{\bar{x}^2 \bar{y} - \bar{x}\bar{x}\bar{y}}{\bar{x}^2 - (\bar{x})^2}.$$

$$\sigma^2 = \frac{n}{n-2} \left(\bar{y}^2 - (\bar{y})^2 - a^2 [\bar{x}^2 - (\bar{x})^2] \right), \quad \sigma_a^2 = \frac{\sigma^2}{n(\bar{x}^2 - (\bar{x})^2)}, \quad \sigma_b^2 = \sigma_a^2 \bar{x}^2.$$



Аппроксимация МНК



Аппроксимация МНК

Некоторые зависимости, можно свести к линейным. Например, $y = e^{ax+b} \Rightarrow \ln y = ax + b$.

Возможно также сведение зависимостей к системам линейных уравнений $A\vec{x} = \vec{b}$, ранг матрицы A должен быть больше количества искомых переменных. Минимизируем $(A\vec{x} - \vec{b})^T(A\vec{x} - \vec{b})$, что приводит к системе уравнений

$$A^T A \vec{x} = A^T \vec{b} \implies \vec{x} = (A^T A)^{-1} A^T \vec{b}.$$

Или $\vec{x} = A^+ \vec{b}$ (псевдообратная матрица), в Octave — «левое деление» $A \backslash b$.



Пример

Пусть заведомо величина изменяется по закону $y = a_0 + a_1 e^{-t} + a_2 t e^{-t}$. В матричном виде $Y = TA$, где T – функциональная матрица, у которой в первом столбце размещены единицы (соответствует умножению на a_0), во втором – функция e^{-t} , а в третьем – $t e^{-t}$. Коэффициенты A найдем при помощи МНК: $A = T \setminus Y$.

```
t = [0 0.3 0.8 1.1 1.6 2.3]';  
y = [0.6 0.67 1.01 1.35 1.47 1.25]';  
T = [ones(size(t)) exp(-t) t.*exp(-t)];  
A = T\y
```



Правило «трех сигм»

При гауссовом распределении случайной величины вероятность

$$P(|x - \bar{x}| < 3\sigma) = 2\Phi(3) = 0.9973.$$

(Φ – нормальное интегральное распределение).

Правило трех сигм: если случайная величина распределена нормально, то абсолютная величина ее отклонения от математического ожидания не превосходит утроенного среднего квадратичного отклонения.

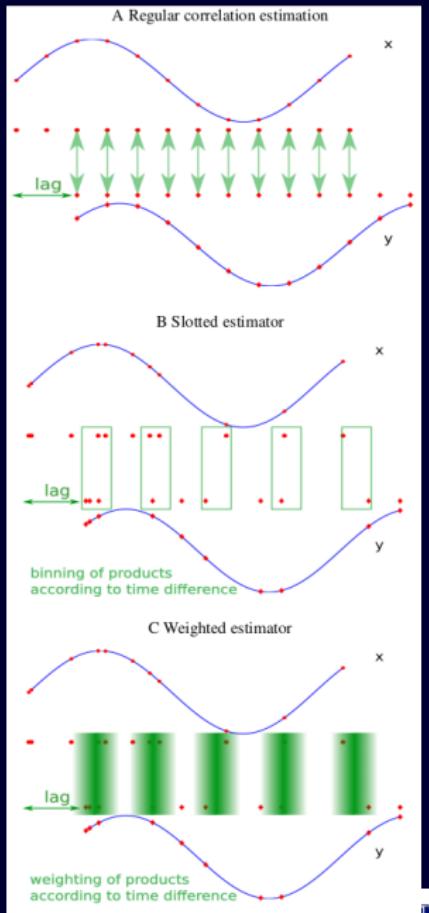
Теорема Ляпунова: случайная величина, являющаяся суммой большого числа взаимно независимых случайных величин, имеет нормальное распределение.



Иррегулярно распределенные данные

БПФ, корреляция, периодограммы и т.п.

- Resampling (если данные достаточно плотно расположены).
- Определение периода как расстояния между минимумами (максимумами) из аппроксимации.
- Auto Regressive Moving Average (ARMA).
- Фильтрация Калмана.
- Метод Ваничека (аппроксимация набора данных рядом синусоид).
- Периодограмма Ломба-Скаргла (ортогонализация пар синусоид введением задержки во времени, Scargle, 1981).
- Irregular Autoregressive Model (IAR).
- Complex IAR (CIAR).



Спасибо за внимание!

[mailto](mailto:)

eddy@sao.ru

edward.emelianoff@gmail.com

