

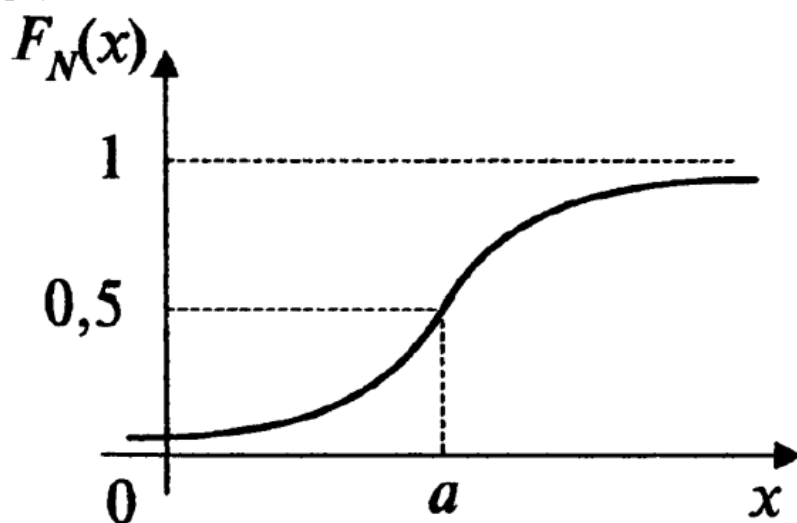
Замечание. Есть некоторая накладка определений:

Есть множество различных распределений. Их логично называть «функциями распределения». Например, функция распределения Гаусса –

$$\varphi_N(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

Все эти функции есть функции плотности вероятностей, но т.к. часто это функции каких-то распределений, логично называть их функция распределений.

Однако «функцией распределения» на ФФ называют не сами функции, а интеграл от них (вероятность того, что величина окажется меньше, чем аргумент):



Я функцией распределения буду называть именно функцию плотности вероятности такого-то распределения. Если речь идёт именно о первообразной, то буду писать «интегральная функция распределения».

Глава 2.

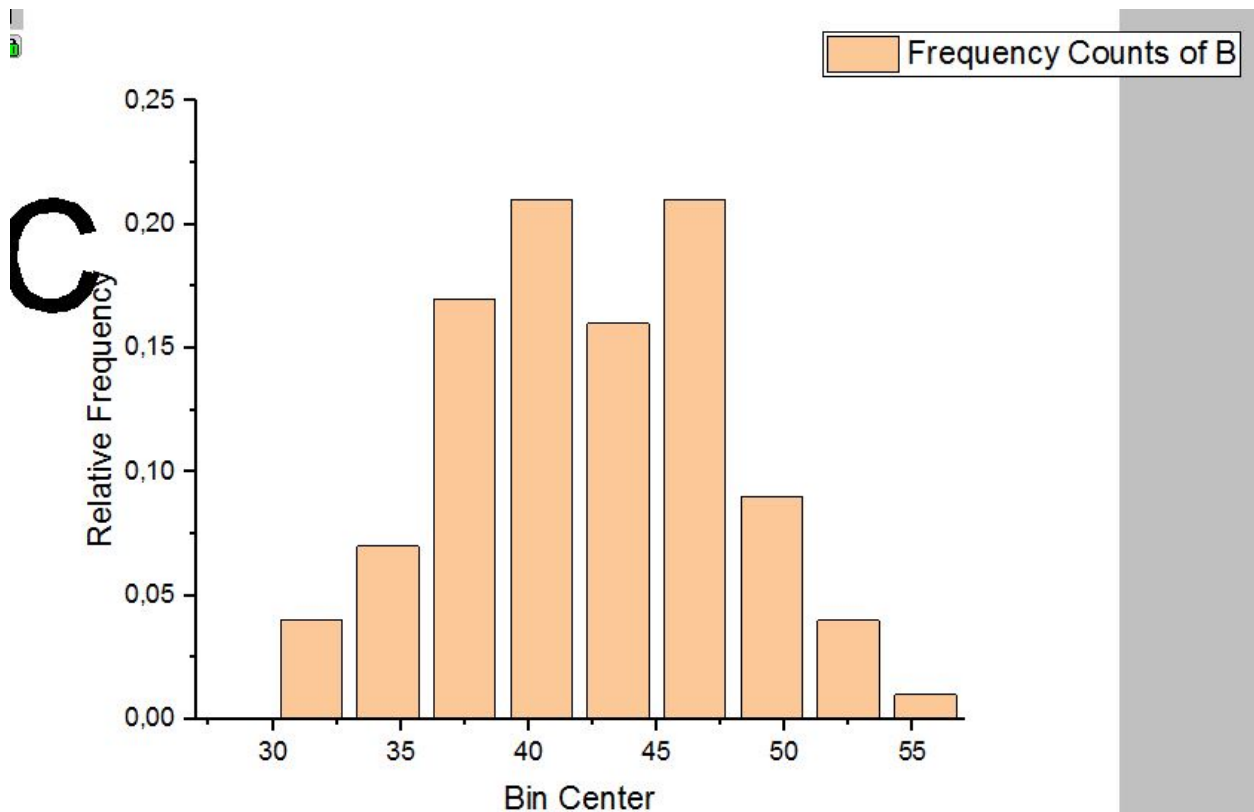
Мы изучили много распределений, и все они зависят от параметров – одного или из нескольких.

Зная параметры, мы всегда можем построить график функции распределения. А можем ли мы решить обратную задачу – по графику функции распределения найти параметры?

Если график известен точно, то да. Например, для нормального распределения вы находите один параметр – a – сразу, как абсциссу максимума, а дисперсию σ^2 – измеряя полуширину.

Однако вся беда, что график плотности вероятности часто получается экспериментально: мы делаем 100500 измерений, набираем статистику, потом строим гистограмму и видим функцию распределения в виде «столбиков», т.е. приближённо.

Например, на легендарном 14 ядерном праче (это где 200 раз Гейгером) у нас получился вот такой график (по оси ординат – вероятности):



Мы ожидаем, что это нормальное распределение, но как нам найти

А) Параметр $\mu = x_0$ (среднее значение) – пик?

Б) Дисперсию σ^2 ?

Сформулируем задачу. У нас есть статистика - последовательность измерений X *конечной* длины n . Для удобства записи её часто записывают в

вектор \vec{X} . Это последовательность – результат измерений одной и той же физической величины, но из-за случайных факторов мы получаем чуть-чуть разные результаты.

Наша цель – *алгоритм*, по которому из статистики (выборки) мы получаем список нужных нас параметров – $\theta_1, \theta_2, \dots, \theta_r$. Для удобства записи его часто

записывают в вектор $\vec{\theta}$.

Пусть имеется случайная выборка $\vec{X} = (X_1, \dots, X_n)$ из генеральной совокупности X , распределение которой $p(x, \vec{\theta})$ известно с точностью до вектора параметров $\vec{\theta} = (\theta_1, \dots, \theta_r)$. Требуется найти оценку параметра $\vec{\theta}$ по случайной выборке \vec{X} .

Есть несколько методов.

Метод максимального правдоподобия.

Идея: $\vec{\theta}$ должен быть такой, что вероятность для этого $\vec{\theta}$ получить выборку \vec{X} больше, чем для каких-либо других $\vec{\theta}$.

Очень логично, не правда ли? Если непонятно – очень рекомендую Райгородского, <https://www.youtube.com/watch?v=ddzt0OkVhfk&list=PLthfp5exSWErTVWq4cVtRXDw5MqBqavJ1&index=4>, с 26:26.

Эта вероятность получить с значением параметром $\vec{\theta}$ выборку \vec{X} традиционно обозначается как $L(\vec{\theta}, \vec{X})$. Тогда условие минимума запишется как

$$\frac{\partial L(\vec{\theta}, \vec{X})}{\partial \theta_k} = 0 \quad \forall \theta_k$$

На практике удобно логарифмировать вероятность:

$$\frac{\partial \ln L(\vec{\theta}, \vec{X})}{\partial \theta_k} = 0 \quad \forall \theta_k$$

Я не сказал, а какой же у этой функции правдоподобия вид. Давайте посмотрим на примере нормальной вероятности.

Вероятность получить значение x при нормальном распределении с матожиданием θ_1 и погрешностью θ_2 есть

$$p(x, \theta_1, \theta_2) = \frac{1}{\theta_2 \sqrt{2\pi}} \exp\left(-\frac{(x - \theta_1)^2}{2\theta_2^2}\right)$$

Если у нас в выборке n значений, то все вероятности перемножаются, тогда функция правдоподобия есть

$$L(\vec{X}; \theta_1, \theta_2) = \frac{1}{(\theta_2 \sqrt{2\pi})^n} \exp\left(-\frac{1}{2\theta_2^2} \sum_{i=1}^n (X_i - \theta_1)^2\right),$$

а её логарифм

$$\ln L(\vec{X}; \theta_1, \theta_2) = -n \ln \sqrt{2\pi} - n \ln \theta_2 - \frac{1}{2\theta_2^2} \sum_{i=1}^n (X_i - \theta_1)^2$$

Система уравнений правдоподобия будет состоять из двух уравнений:

$$\begin{cases} \frac{\partial \ln L}{\partial \theta_1} = \frac{1}{\theta_2^2} \sum_{i=1}^n (X_i - \theta_1)^2 = 0, \\ \frac{\partial \ln L}{\partial \theta_2} = -\frac{n}{\theta_2} + \frac{1}{\theta_2^3} \sum_{i=1}^n (X_i - \theta_1)^2 = 0. \end{cases}$$

Решение системы

$$\hat{\theta}_1 = \frac{1}{n} \sum_{i=1}^n X_i, \quad \hat{\theta}_2^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

В главе 2 у нас была последовательность X , где мы многократно измеряли одну и ту же величину.

Меж тем частенько мы измеряем что-то, меняя параметры. Например, мы предполагаем линейную зависимость $y=kx+b$, и при наборе X измеряем игреки.

Как нам восстановить по полученным точкам k и b ?

Метод моментов.

Идея: мы считаем экспериментально моменты:

матожидание $\langle x \rangle$

средний квадрат $\langle x^2 \rangle$

средний куб $\langle x^3 \rangle$

и т.д.

И считаем, что они у нас абсолютно точные. Тогда мы задаёмся вопросом – а при каких значениях параметров $\theta_1 \dots \theta_n$ мы бы получили такие моменты?

Решаем систему из n уравнений.

Пример.

Возьмём скромную выборку из 5 значений:

2, 0, 1, 0, -2.

Известно, что СВ подчиняется распределению

$$N(A, B, C) \frac{A}{1 + C(x - B)^2}, \text{ где } N(A, B, C) = \frac{1}{\int_{-\infty}^{\infty} \frac{A}{1 + C(z - B)^2} dz}$$

Найти три параметра – A, B, C .

Решение: считаем моменты:

$$\langle x \rangle = \frac{1}{5} (2 + 0 + 1 + 0 + (-2)) = \frac{1}{5}$$

$$\langle x^2 \rangle = \frac{1}{5} (4 + 0 + 1 + 0 + 4) = \frac{9}{5}$$

$$\langle x^3 \rangle = \frac{1}{5} (8 + 0 + 1 + 0 + (-8)) = \frac{1}{5}$$

Зададимся вопросом: при каких значениях A, B, C $\langle x \rangle$ окажется $\frac{1}{5}$, $\langle x^2 \rangle$ окажется равным $\frac{9}{5}$, а $\langle x^3 \rangle$ окажется равным $\frac{1}{5}$?

Да при таких, что

$$\begin{cases} \int_{-\infty}^{\infty} \frac{A}{1 + C(x - B)^2} N(A, B, C) x = \frac{1}{5} \\ \int_{-\infty}^{\infty} \frac{A}{1 + C(x - B)^2} N(A, B, C) x^2 = \frac{9}{5} \\ \int_{-\infty}^{\infty} \frac{A}{1 + C(x - B)^2} N(A, B, C) x^3 = \frac{1}{5} \end{cases}$$

Осталось взять интегралы, что легко делается с помощью Вольфрама, и решить систему.

Метод наименьших квадратов

Служит уже для другой цели – теперь мы не меняем параметр x , измеряя при этом y .

Не спешите закрывать методичку, думая «а я слушал Митина и всё про МНК знаю». Нет, будет и новое: вам, вероятно, он известен для аппроксимации $y=kx+b$. На самом деле теоретическая зависимость y от x может быть произвольной.

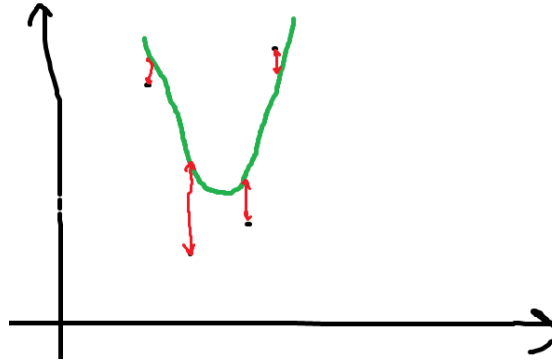
$$Y = f(X) + \varepsilon(X), \quad (3.1)$$

где $f(X)$ – скалярная функция, $\varepsilon(X)$ – случайная ошибка, то есть случайная составляющая, порожденная либо действием физических флуктуаций, либо случайными ошибками измерений величины $f(X)$, либо и тем и другим одновременно.

Будем считать, что для каждого X математическое ожидание $\varepsilon(X)$ равно нулю, то есть отсутствует систематическая погрешность модели.

$f(x)$ у нас будет иметь вид $\varphi(X; \vec{c})$, т.е. это функция от X , но зависящая от списка параметров (вектора параметров) \vec{c} . Наша цель – подобрать \vec{c} так, чтобы эта функция как можно лучше аппроксимировала экспериментальную. Например, в случае $y=kx+b$ в \vec{c} входят k и b .

Нужно минимизировать вот такую величину. Это сумма квадратов длин



красных стрелочек. Чем она меньше, тем лучше аппроксимация!

$$\sqrt{\sum_{i=1}^n \rho(X_i) (Y_i - \varphi(X_i; \vec{c}))^2}$$

, т.е. найти такой \vec{c} , чтобы эта

величина была минимальной.

$\rho(X_i)$ – вес i -того измерения. По умолчанию 1, но если вы хотите сделать иным – пожалуйста, делайте.

Или вместо максимизирования квадратного корня можно максимизировать подкоренное выражение:

$$\sum_{i=1}^n \rho(X_i) (Y_i - \varphi(X_i; \vec{c}))^2 = \min$$

В общем случае решение этой задачи очень сложно. Чаще рассматривают

$$\varphi(x; \vec{c}) = \sum_{j=1}^p c_j \varphi_j(x)$$

частный случай, когда параметры линейно.

зависит от

Это не только $y=kx+b$, это и $y=C_0+C_1*x+C_2*x^2+C_3*x^3\dots$, и $y=C_0+C_1*\exp(x)+C_2*\sin(x)+C_3*\ln(x)\dots$

Подставим вид φ в том, что надо минимизировать:

$$\begin{aligned} \|Y - \varphi(X; \vec{c})\|_{L_2}^2 &= \sum_{i=1}^n \rho(X_i) \{Y_i^2 - 2Y_i \sum_{j=1}^p c_j \varphi_j(X_i) + \sum_{j=1}^p \sum_{k=1}^p c_j c_k \varphi_j(X_i) \varphi_k(X_i)\} \\ &= (Y, Y) - 2 \sum_{j=1}^p c_j (Y, \varphi_j) + \sum_{j=1}^p \sum_{k=1}^p c_j c_k (\varphi_j, \varphi_k) \\ &= \min, \end{aligned} \tag{3.6}$$

где $(Y, Y) = \sum_{i=1}^n \rho(X_i) Y_i^2$, $(Y, \varphi_j) = \sum_{i=1}^n \rho(X_i) Y_i \varphi_j(X_i)$ и

$(\varphi_j, \varphi_k) = \sum_{i=1}^n \rho(X_i) \varphi_j(X_i) \varphi_k(X_i)$ – скалярные произведения. Запишем теперь

условия минимума квадрата нормы разности (3.6):

$$\frac{\partial}{\partial c_j} \|Y - \varphi(X; \vec{c})\|_{L_2}^2 = 2\{-(Y, \varphi_j) + \sum_{k=1}^p c_k(\varphi_j, \varphi_k)\} = 0, \quad j = 1, \dots, p \quad (3.7)$$

или

$$\sum_{k=1}^p c_k(\varphi_j, \varphi_k) = (Y, \varphi_j), \quad j = 1, \dots, p. \quad (3.8)$$

Заметим, что в случае, функции $\{\varphi_j(x)\}$, $j = 1, \dots, p$ не просто линейно независимы, а ортогональны или даже ортонормированы, то есть $(\varphi_j, \varphi_k) = \delta_{ij}$, уравнения (3.7) сильно упрощаются:

$$c_j = (Y, \varphi_j), \quad j = 1, \dots, p. \quad (3.9)$$

Именно поэтому для построения наилучшего среднеквадратичного приближения рекомендуется использовать системы ортогональных полиномов.

Замечание: условие ортогональности здесь не «интеграл от а до b от произведения функций есть 0», а «сумма по i от 1 до n $\varphi_j(x_i) \cdot \varphi_k(x_i)$ равна 0». Так что не спешите расчехлять ваши любимые КОПы, у них условие ортогональности другое.

Тем более что часто, как родителей не выбирают, не выбирают и $\varphi_j(x_i)$ – они обусловлены тем, какая у вас теоретическая зависимость. Например, в случае

$$y = c_0 + c_1 x + c_2 x^2 + c_3 x^3 \dots \quad \varphi_j(x) = x^j, \quad j = 0, 1, \dots, p.$$

Так что в общем случае надо решать СЛАУ

$$\sum_{k=1}^p c_k(\varphi_j, \varphi_k) = (Y, \varphi_j), \quad j = 1, \dots, p$$

Пример. У нас вот такая серия измерений

X	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0
Y	7.0	8.0	3.0	4.0	5.0	6.0	7.0	9.0	6.0	7.0

Погрешности всех измерений равны, поэтому вес – 1.

Аппроксимировать будем полиномом степени p $\sum_{j=0}^p c_j x^j$, число параметров – p+1.

Нужно минимизировать вот такую величину

$$\sum_{i=1}^{10} \left(y_i - \sum_{j=0}^p c_j x_i^j \right)^2$$

А далее составители методички прекратили решать тот пример 😞. Понятно, что нужно записать СЛАУ и решить её.