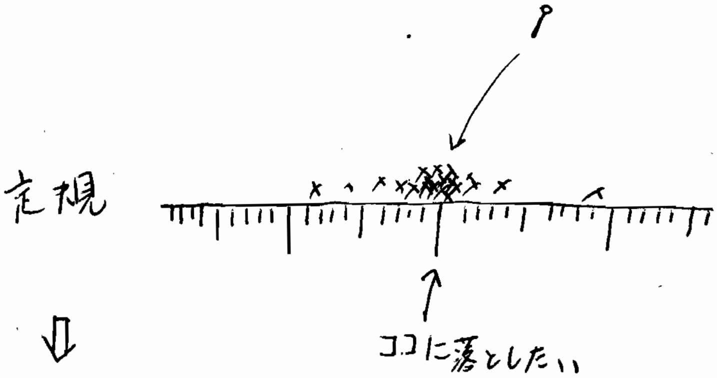
- 測定誤差の数学的取扱と最小二乗法（フィット）

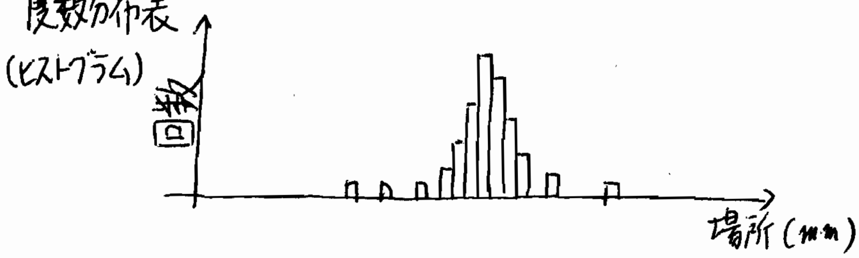
→ 中心極限定理と確率分布

→ データ点に対してどのように理論の線を引くか
（物理学実験では使わないが、誰がやっても同じ線を引く理屈）

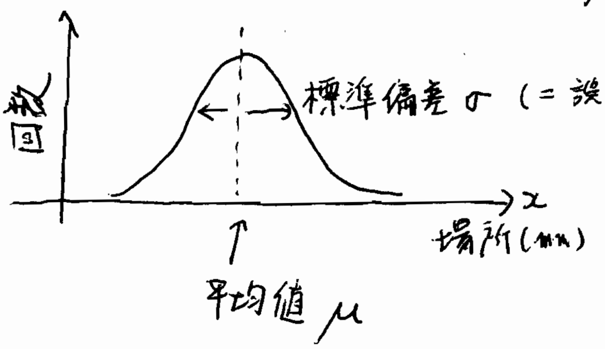
☆ 的を狙ってピンを落とす動作をくり返すと‥‥



⇩



偶然誤差によってちらばる．このとき、正規分布に近くなることが多い．



⇒ $\boxed{\dfrac{1}{\sqrt{2\pi}\sigma}\exp\left(-\dfrac{(x-\mu)^2}{2\sigma^2}\right)}$
（正規分布の式）

正規分布‥‥・狙った値が分布の平均値になる
・平均値の頻度（回数）が最も大きく、離れると小さくなる

☆ どのような測定値の偶然誤差や、分布も正規分布になることが多い．
・経験的直観的には、当然！
・数学（統計学）的にも、強い根拠がある．

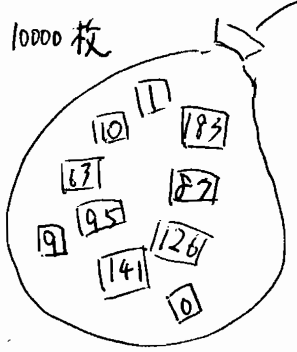
☆ 中心極限定理

……ある平均と分散(標準偏差)をもついかなる分布においても(μ, σ)，大きさ n の無作為標本に基づく標本平均 $\bar{x}$ は平均 μ，標準偏差 $\frac{\sigma}{\sqrt{n}}$ の正規分布に従う。

⇩ かみくだいて言うと

どんな測定値の分布(偶然誤差の分布)も正規分布になる！

0-200のカード
10000枚



20枚(n枚)とることをくりかえす

		平均値
1回目	5 21 … 11 → 平均をとる	μ_1
2回目	9 63 … 191 →	μ_2
⋮	⋮ →	⋮

このμᵢの分布は正規分布になる

正規分布ではない(一様分布)のカードの組 ⇒ 測定値の分布は正規分布になる

シミュレーション：http://www.kwansei.ac.jp/hs/z90010/sugakuc/toukei/tyuusin2/chuusin.htm

シミュレーションで もとの分布が
- 正規分布
- 一様分布
- 指数分布

の場合でも，平均値の分布は正規分布になる。

☆ 実験的には中心極限定理の意味あいは

・実験システム(実験装置・実験環境・測定されるそのもの…)が，それぞれ正規分布とは限らない測定結果のふれ(誤差)を発生させていたとしても，何度も測定した結果の平均値とそのふれは正規分布になる。

誤差が誤差どうしで影響しあう場合(共働現象)などはこの限りではないが，ほとんどの場合，正規分布でよい近似となっている。

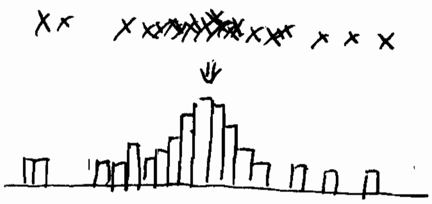
> なぜ誤差論で正規分布にこだわるかというと，数学的取扱いの簡単さと中心極限定理に理由がある。

☆ データ点と理論式(理論直線・曲線)をうまくあわせる方法

・まず、平均値 とは何だろう?

1次元の分布

の度数分布



を考える. これは正規分布をしていると仮定しよう.

この正規分布の中心(平均)値 μ は データの平均値

分散 σ^2 は 〃 分散 $\sigma^2 = \sum_i (x_i - \mu)^2$

標準偏差 σ は $\sqrt{\sigma^2} = \sqrt{\sum_i (x_i - \mu)^2}$

逆に考える → 平均値とは、どのデータ点からも近い点のこと

$$\prod_i \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right) dx_i$$

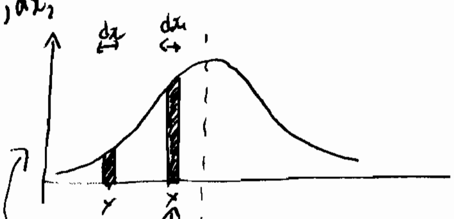
$$= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_1-\mu)^2}{2\sigma^2}\right) dx_1 \cdot \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_2-\mu)^2}{2\sigma^2}\right) dx_2 \cdots$$

$$= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\sum_i (x_i-\mu)^2}{2\sigma^2}\right) dx_i$$

$(x_i-\mu)^2$ がそれぞれの確率みたいなものになる.

⇒ $\sum_i (x_i - \mu)^2$ を最小にすると確率が最大になる

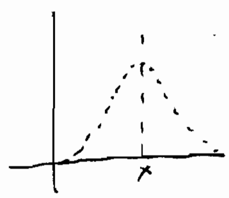
(※)



こっちの方がより近い.

度数を確率と考えよう⇔確率分布と考える.

データ1点の場合



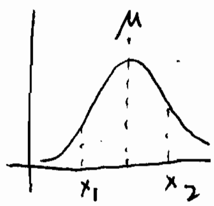
データ = 平均値 = 中心値

$(x_1 - \mu)^2$ を最小

⇔ $x_1 = \mu$

$\boxed{\mu を動かす}$

2点の場合



$\sigma^2 = \sum (x_i - \mu)^2$

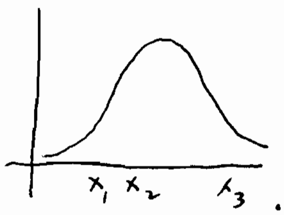
$(x_1-\mu)^2 + (x_2-\mu)^2$

μ でビブンして

$(\sigma^2)' = 4\mu - 2x_1 - 2x_2$

$\mu = \dfrac{x_1 + x_2}{2}$

3点の場合



同じように計算して

$\mu = \dfrac{x_1 + x_2 + x_3}{3}$

一般に $\mu = \dfrac{\sum^n x_i}{n}$

Maximum Likelihood Method
=最大尤(ゆう)度法 =「もっともありそうな」法

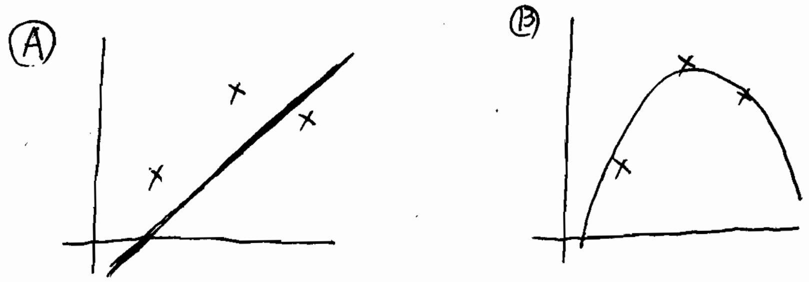
最大確率で考えても正しく平均値が算出される.
当然分散も計算され、正規分布がもとまる.

☆データ点と理論をあわせる。

理論は所与（はじめにあたえられている）とする

→直線の場合もあり、曲線の場合もある。

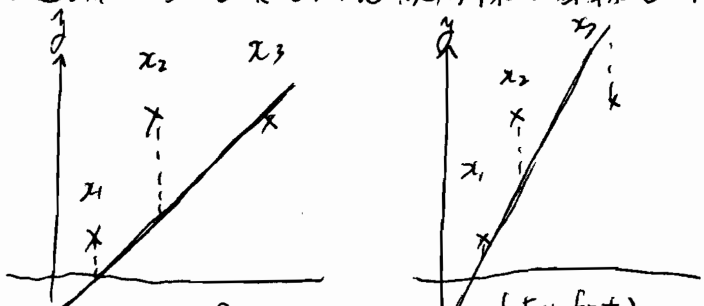
→切片がきまっている場合もある



点の位置は同じだが、理論（Ⓐ直線Ⓑ放物線）が与えられており
どちらもありうる。

☆最小二乗法（MS Excelなどの回帰計算と同じ）

直線理論を考える（もちろん放物線や曲線も可能）

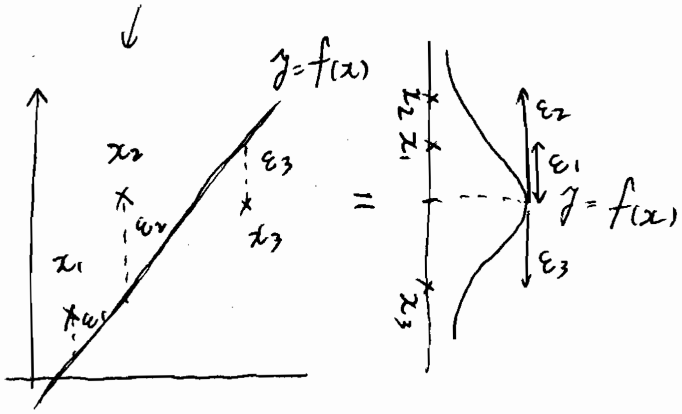


（へたくそfit）　（よいfit）

よいfitは直線とデータ点のy軸での距離が小さい。

⇓

さっきやった一次元の考え方を少しだけ拡張する



$y=f(x)=ax+b$

$y=f(x)$を$y=0$（もしくは$y=b$）にしてやるとさっきやった一次元の場合と全く同じ。

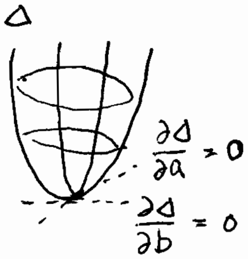
$\varepsilon_i = y_i - a x_i - b$ で $\sum_i^n \varepsilon_i^2$ を最小にする $y=\underset{\sim}{a}x+\underset{\sim}{b}$ の(a, b)を求めればよい。

このとき一番確率が高い（ありそうな）直線となる。

・計算（テキストP20付近）

$$\sum \varepsilon_i^2 = \sum (y_i - a x_i - b)^2 = \Delta(a,b)$$ とおく

aとbを変数と考えてこの式をaとbで偏微分して0になる点がΔ(a,b)を最小となる点.



Ⓐ $\dfrac{\partial \Delta(a,b)}{\partial a} = 0 \Leftrightarrow \sum_i^n \cancel{-2}\, x_i (y_i - a x_i - b) = 0$

Ⓑ $\dfrac{\partial \Delta(a,b)}{\partial b} = 0 \Leftrightarrow \sum_i^n \cancel{-2}\, (y_i - a x_i - b) = 0$

・両方を解く

Ⓐ $\sum_i^n (x_i y_i - a x_i^2 - b x_i) = 0$　　$\sum_i^n x_i = n\langle x\rangle$, $\langle x\rangle$はx_iの平均値 とおくと

$\langle x^2\rangle a + \langle x\rangle b = \langle xy\rangle$ …… ①

Ⓑ $\sum^n (y_i - a x_i - b) = 0$

$\langle x\rangle a + b = \langle y\rangle$ …… ②

結局 平均値 $\langle x\rangle$, $\langle y\rangle$, $\langle xy\rangle$, $\langle x^2\rangle$ を求めれば, (a, b)が出てくる.

$$a = \frac{\langle xy\rangle - \langle x\rangle\langle y\rangle}{\langle x^2\rangle - \langle x\rangle^2}, \quad b = \frac{\langle x^2\rangle\langle y\rangle - \langle x\rangle\langle xy\rangle}{\langle x^2\rangle - \langle x\rangle^2}$$

・では問題をやってみよう

1の答えは $\Delta L_0 + At$

2, ΔLは 1: $(L_1 - L_1)$, 2: $(L_2 - L_1)$, 3: $(L_3 - L_1)$ を入れて下さい
（小数点以下2ケタ）

グラフの範囲は $0 < t < 100$℃, $-0.20 < \Delta L < 0.60$ mm かそれ以上に.
・目分量でデータ点に直線を点線で引きましょう

3. 計算のケタは最大限とりましょう.

対応は $t \to x$, $\Delta L_i \to y$, $a \to A$, $b \to \Delta L_0$ です.

最後の ΔL_0, A の値は -0.110 mm, 6.51×10^{-3} mm℃$^{-1}$ になります.
・この数字を使って最小二乗法による直線を実線で引きましょう

4. $L_0 = L_1 + \Delta L_0 =$ ……

$\therefore \alpha = \dfrac{A}{L_0} =$ ——— $=$ 　　　となります.

補遺

- 求め方からわかるように $y=f(x)$ は好きな関数が使えます。

- 最小二乗法は誤差が同じであると仮定しているので
それぞれのデータ点の誤差が異なるときは最小二乗法は使えません。
これはMS-Excelの回帰計算でも同じです。
このような場合は $\sum_i \frac{\varepsilon_i^2}{\sigma_i^2}$ として、誤差を σ_i としたカイ2乗フィットを
おこないます。

- 誤差の分布が正規分布でな ときですら、確率分布がわかっていれば
3ページの(※1)にもどって Maximum Likelihood Method でフィットできます。

- 実験のデータ点は目でみて線を引けばよく、
最小二乗法で計算する必要はありません。

来週(5月21日)の実験は 13:00 から南棟2階
物理第1実験室で「重力加速度の測定」を行います。
テキスト35ページから読んできて下さい。