Uncertainity in Cyber Infrastructure

GEON - cyber infrastructure in the geosciences - example.

Ultimate objectives:

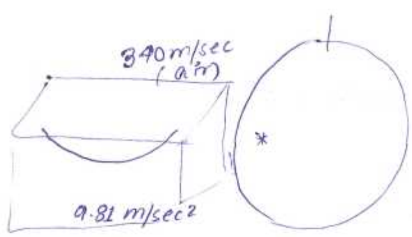
- earthquake prediction & analysis.
- look for minerals. (metal ores, oil)
  water
- pollution.

We want to know what is down there (1-2-5 Km)

- digging a well.
- direct measurements are expensive so you need indirect ones.

Geo Measurements:

* satellite photos. - NASA.
* collect samples, do analysis.



* active seismic data from the seismic stations. - seismic stations.
* active seismic data. - at universities
* gravity measurements. - at univ, oil companies, at NASA
  (Earth based and satellite based).

How did scientists operate?
Old ways network.

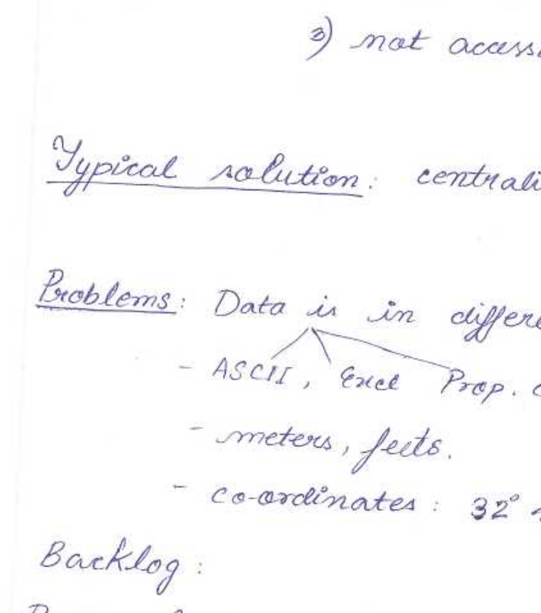
Disadvantage: 1) Difficult to get.
2) slow
3) not accessible to everybody.

Typical solution: centralisation.

Problems: Data is in different formats.
- ASCII, Excel Prop. databases.
- meters, feets.
- co-ordinates: 32° 17' ...

Backlog:

Better Solution: To keep every piece of data in its place.



Cyber Infrastructure:

Uncertainity in general.

what computers do? process numbers.

where do numbers come from:

- measurements.
- from an expert.

this information is never exact.

confusing → error (diff meanings)
- meaning: true bad mistake.
- small inaccuracy. (measurement error, approximation error).

x - actual (unknown) value.

$\tilde{x}$ - estimate from measurement or from an expert.

$\Delta x \overset{def}{=} \underbrace{\tilde{x}}_{\text{known}} - \underbrace{x}_{\text{not known}} \Rightarrow$ estimation error, approxn. error, measurement error.

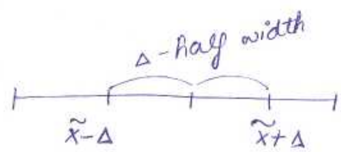
<u>usually</u>: we know the upper bound Δ

$|\Delta x| \leq \Delta$.

$\tilde{x} = 312$ Kg.

$\Delta = 1$ Kg.

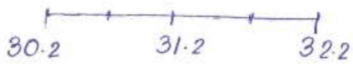
$x = \tilde{x} - \Delta x$



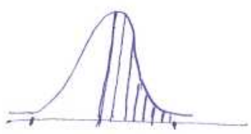
$x \in [\tilde{x}-\Delta, \tilde{x}+\Delta]$ → interval.

interval uncertainity;

$[\underline{x}, \overline{x}]$ → upper bound, → lower bound



Some values are more frequent and some are less frequent.



bell-shaped,

Gaussian Distribution.

Probability density.

$$\rho = \frac{\Delta P}{\Delta x}$$

$$\rho(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right)$$

a, σ – how do we determine them.

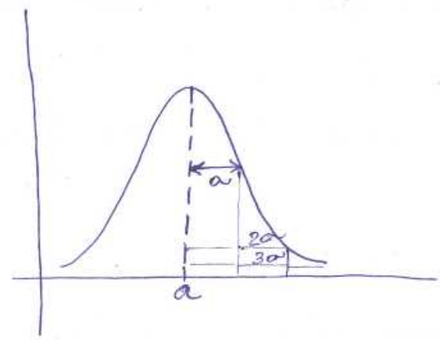
$$a = \frac{x_1 + x_2 + \cdots x_n}{n}$$

$$\sigma = \sqrt{\frac{(x_1-a)^2 + \cdots + (x_n-a)^2}{n}}$$

HW: (1) Simple code for simulating Gaussian distribution.

(2) Run it several times

(3) Plot it & get the graphical form



$x = a, \rho(x) = \frac{1}{\sqrt{2\pi}\sigma}$

$x = a+\sigma, \rho(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2})$

$\frac{1}{e^{1/2}} = \frac{1}{\sqrt{e}}$, $e = 2.71828$

$\simeq \frac{1}{1.6} \simeq 0.6$

$x = a+2\sigma, \rho(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-2)$

$1/e^2 = 1/9 = 0.1$.

$x = a+3\sigma, \frac{1}{e^4 \cdot e^{0.5}} = \frac{1}{64 \cdot 1.6} \simeq \frac{1}{100}$

Central Limit Theorem:

uniform on $[0,1]$

$0 \longleftrightarrow 1 \leftarrow$ (provided by compiler)

easiet way:

random() − 0.5 // so that mean is 0

gauss_first = 0.0;

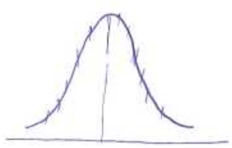
for i = 1:12

{gauss = gauss + (random() − 0.5); }

$\rho(x) = 1.$

$$\int_{-0.5}^{0.5} (x-0)^2 dx = \frac{x^3}{3}\Big|_{-\frac{1}{2}}^{\frac{1}{2}} = \frac{1}{8\cdot 3} - \left(\frac{-1}{8\cdot 3}\right)$$

Maximum likelihood: Method:



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

$$= \frac{1}{\sqrt{2\pi}\cdot\sigma} \exp\left(-\frac{(x_n-a)^2}{2\sigma^2}\right) \to \max \quad a, \sigma$$

$$\ln(\exp(z)) = z.$$

$$\ln \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_n-a)^2}{2\sigma^2}\right)$$

We had: reminder of probability things:

Plan: - how to get estimates.

- how to process uncertainities.

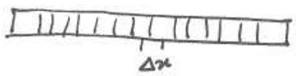
- how to deal with interval uncertainity.

$\rho(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right)$ - probability density for Gaussian distribution.

(random variable.)

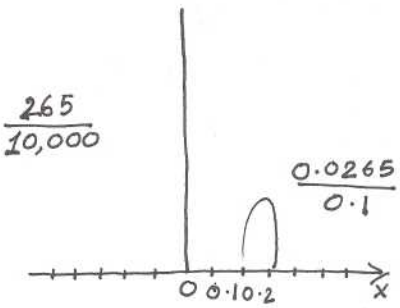
$\text{Prob}\,(x \leq \delta \leq x+\Delta x) \simeq \rho(x).\Delta(x).$

$\rho(x) \approx \frac{\text{Prob}\,(x \leq \delta \leq x+\Delta x)}{\Delta x}$



$\frac{265}{10{,}000}$

$\frac{0.0265}{0.1} \simeq 0.265.$



$x_1\ x_2 \ldots\ldots x_n.$

so far we know, a, σ

us simulated distribution.

we predicted probability for distribution

Question:

how do we find a and σ?

Given: $x_1, \ldots\ldots x_n$.

we need to estimate a and σ.

select most probable values, a, σ.

H_1: $a = 0$, $\sigma = 1$. $\quad \frac{1}{\sqrt{2\pi}}.\exp(-0^2).$ $\Rightarrow$ more probable.

H_2: $a = 20$, $\sigma = 1$. $\quad \frac{1}{\sqrt{2\pi}} \exp\left(\frac{20^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi}} \exp(-200).$

$x_1 = 0$

$\frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x_1-a)^2}{2\sigma^2}\right).\Delta x.$ (ignore: we are comparing diff probabilities).

Prob = p.density × Δx.

$$\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_2-a)^2}{2\sigma^2}\right)$$

$p(\text{rain}) = 50\%$.
$p(O) = 55\%$
$p.(\text{rain}) * p(O) \rightarrow$ independent events.

. . . .

$$\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_n-a)^2}{2\sigma^2}\right).$$

$$\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(-x_1-a)^2}{2\sigma^2}\right) . \Delta x . \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(-x_2-a)^2}{2\sigma^2}\right) . \Delta x . \cdots \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(-x_n-a)^2}{2\sigma^2}\right) . \Delta x$$

$\rightarrow$ max. a, σ.

Tricks: $a^b . a^c = a^{b+c}$.

$$\underbrace{a \ldots a}_{b} \; \underbrace{a \ldots a}_{c} = a^{b+c}.$$

$$\frac{\Delta x^n}{\sqrt{2\pi}^n} \times \frac{1}{\sigma^n} \times \exp\left(-\frac{(x_1-a)^2}{2\sigma^2}\right) . \exp\left(-\frac{(x_2-a)^2}{2\sigma^2}\right) \cdots \exp\left(-\frac{(x_n-a)^2}{2\sigma^2}\right)$$

$$= \text{const} . \frac{1}{\sigma^n} . \exp\left(-\frac{(x_1-a)^2}{2\sigma^2}\right) . \exp\left(-\frac{(x_2-a)^2}{2\sigma^2}\right) \cdots \exp\left(-\frac{(x_n-a)^2}{2\sigma^2}\right) \rightarrow \text{max. no.}$$

$$= \text{const} . \frac{1}{\sigma^n} . \exp\left(-\frac{(x_1-a)^2}{2\sigma^2} - \frac{(x_2-a)^2}{2\sigma^2} - \cdots - \frac{(x_n-a)^2}{2\sigma^2}\right)$$

$$-\frac{(x_1-a)^2}{2\sigma^2} - \frac{(x_2-a)^2}{2\sigma^2} \cdots - \frac{(x_n-a)^2}{2\sigma^2} \Big] \rightarrow \text{max } a, \sigma.$$

$$\frac{(x_1-a)^2}{2\sigma^2} + \frac{(x_2-a)^2}{2\sigma^2} + \cdots + \frac{(x_n-a)^2}{2\sigma^2} \rightarrow \text{min } a, \sigma.$$

Differentiate by a, equate the derivative to 0.

$$. \frac{dx}{da} . \frac{(x_1-a)^2}{2\sigma^2} + \cdots$$

$$= \frac{1}{2\sigma^2} . \left[(x_1-a)^2 + (x_2-a)^2 + \cdots (x_n-a)^2\right]$$

$$= \frac{1}{2\sigma^2} . \left[2(x_1-a) . \left(\frac{dx_1}{da} - 1\right) + 2(x_2-a) . \left(\frac{dx_2}{da} - 1\right) \cdots + 2(x_n-a) . \left(\frac{dx_n}{da} - 1\right)\right.$$

$= \cancel{\frac{1}{2a^2}} \cdot 2\cancel{(x_1 - a)}.$

$= \frac{1}{2a^2} \cdot \left[2(x_1 - a) \cdot (-1) + 2(x_2 - a) \cdot (-1) + \cdots \quad + 2(x_n - a) \cdot (-1) \right]$

$= \frac{1}{2a^2} \left[2(a - x_1) + 2(a - x_2) + \cdots \quad + 2(a - x_n) \right] = 0$

$a - x_1 + a - x_2 + \cdots \quad a - x_n = 0$

$na = x_1 + x_2 + \cdots x_n$

$\frac{x_1 + x_2 + \cdots x_n}{n} = a \text{ (Mean)}.$

$(x^{-n})' = nx^{-n-1}$

$\frac{1}{a^2} = a^{-2}$

Finding Sigma;

$-\frac{(x_1 - a)^2}{2a^2} - (\cdots) \cdots - \frac{(x_n - a)^2}{2a^2} - n \ln a \rightarrow \max a, a.$

$\frac{d}{da}(a^{-2}) = a^{-3}$

$a.b$

$\frac{d}{dc}(a \cdot b) = \frac{da}{dc} \cdot b + \frac{db}{dc} \cdot a$

$\frac{(x_1 - a)^2}{2a^2} + \cdots \quad + \frac{(x_n - a)^2}{2a^2} + n \ln a$

$\cancel{2(x_1 - a)} \cdot \frac{1}{2a^2} \cdot \left[(x_1 - a)^2 + \cdots + (x_n - a)^2 \right] + n \ln a.$

$= \frac{-2}{2} \cdot a^{-3} \cdot \left[(x_1 - a)^2 + \cdots + (x_n - a)^2 \right] + \frac{1}{2a^2} \times 0 \qquad + n \frac{1}{a} = 0$

$\cancel{\frac{1}{\sqrt{2\pi}}} \cancel{\pm n} \cancel{\frac{da}{a^2}} \cdot \left[(x_1 - a)^2 + (x_n - a)^2 \right] + \frac{n}{a} = 0$

$= -\cancel{2} \frac{1}{\cancel{2} a^3} \left[(x_1 - a)^2 + \cdot (x_n - a)^2 \right] = \frac{n}{a}$

$a^2 = \cancel{\tfrac{1}{2}} \left[\frac{(x_1 - a)^2 + \cdot (x_n - a)^2}{n} \right]$ Standard Deviation.

Data Fusion:

Several sources of information are merged together to form data fusion.

x - same quantity

$\tilde{x}_1, \tilde{x}_2$ - two results of measuring the same quantity

$\Delta x_1 = \tilde{x}_1 - x$ ÷ normally distributed with 0 mean & S.D σ_1

$\Delta x_2 = \tilde{x}_2 - x$ ÷ " " " " " & S.D σ_2.

$\tilde{x}_1 = 140 \quad \pm 30 = \sigma_1$

$\tilde{x}_2 = 110 \quad \pm 10 = \sigma_2$

common sense approach:

Choose a more accurate value

Ideally we should not ignore, we should combine.

Idea: Same (v.e) idea:

$$\frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(x-\tilde{x}_1)^2}{2\sigma^2}\right) \times \Delta x \cdot \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{(x-\tilde{x}_2)^2}{2}\right) \times \Delta x = \text{Prob. of a given value } x$$

$\Rightarrow$ max.

in first measurement we get x_1 & in second we get x_2.

$$-\frac{(x-\tilde{x}_1)^2}{2\sigma_1^2} - \frac{(x-\tilde{x}_2)}{2\sigma_2^2} \Rightarrow max.$$

$$\Rightarrow \frac{(x-\tilde{x}_1)^2}{2\sigma_1^2} + \frac{(x-\tilde{x}_2)^2}{2\sigma^2} \Rightarrow min. \text{ (least squares method) (Gauss).}$$

$$\frac{2(x-\tilde{x}_1)}{2\sigma_1^2} + \frac{2(x-\tilde{x}_2)^2}{2\sigma_2^2} = 0$$

or $x(\sigma_1^{-2} + \sigma_2^{-2}) = \tilde{x}_1 \sigma_1^{-2} + \tilde{x}_2 \sigma_2^{-2}$

or $x = \dfrac{\tilde{x}_1 \sigma_1^{-2} + \tilde{x}_2 \sigma_2^{-2}}{\sigma_1^{-2} + \sigma_2^{-2}}$ (combined fields estimate)

Case 1:

$\tilde{x}_1 = 140, \quad \pm 30 = \sigma_1$

$\tilde{x}_2 = 110, \quad \pm 10 = \sigma_2$

$$= \frac{110 \cdot \frac{\cancel{1}}{\cancel{100}} + 140 \cdot \frac{\cancel{1}}{\cancel{100}}}{\frac{\cancel{1}}{\cancel{100}} + \frac{\cancel{1}}{\cancel{100}}} = \frac{110+140}{2} = 125.$$

$$\frac{140 \cdot \frac{1}{900} + 1\cancel{4}0 \cdot \frac{1}{100}}{\frac{1}{900} + \frac{1}{100}} = \left[\frac{140 \cdot \frac{1}{9} + 1\cancel{4}0}{\frac{1}{9}+1} = \frac{\frac{140 + 9 \cdot 1\cancel{4}0}{9}}{\frac{10}{9}} = \frac{140 + 990}{10} = 113 \right]$$

$$\boxed{\begin{array}{l} X = \tilde{x}_1 \cdots \sigma_1 \\ X = \tilde{x}_2 \cdots \sigma_2 \\ \vdots \\ X = \tilde{x}_n \cdots \sigma_n \end{array}}$$

$$\frac{1}{\sqrt{2\pi}\,\sigma_1} \exp\left(\frac{-(x-\tilde{x}_1)}{2\sigma_1^2}\right)^2 \cdot \Delta x \cdot \frac{1}{\sqrt{2\pi}\,\sigma_2} \cdot \exp\left(-\frac{(x-\tilde{x}_2)}{2\sigma_2^2}\right)^2 \cdots$$

$$\frac{1}{\sqrt{2\pi}\,\sigma_n} \exp\left(-\frac{(x-\tilde{x}_n)}{2\sigma_n^2}\right)^2$$

$$= -\left(\frac{x - \tilde{x}_1}{2\sigma_1^2}\right)^2 - \left(\frac{x - \tilde{x}_2}{2\sigma_2^2}\right)^2 \cdots - \left(\frac{x-\tilde{x}_n}{2\sigma_n^2}\right)^2 \Rightarrow \max$$

$$= \left(\frac{x - \tilde{x}_1}{2\sigma_1^2}\right)^2 + \left(\frac{x - x_2}{2\sigma_2^2}\right)^2 \cdots + \frac{(x-\tilde{x}_n)^2}{2\sigma_n^2} \Rightarrow \min$$

$$\frac{2(x - \tilde{x}_1)}{2\sigma_1^2} + \frac{2(x - \tilde{x}_2)}{2\sigma_2^2} + \cdots + 2\frac{(x-\tilde{x}_n)}{2\sigma_n^2} = 0$$

$$\frac{(x - \tilde{x}_1)}{\sigma_1^2} + \frac{(x - \tilde{x}_2)}{\sigma_2^2} + \cdots \frac{(x - \tilde{x}_n)}{\sigma_n^2} = 0$$

$$x\,(\sigma_1^{-2}+\sigma_2^{-2}+\cdots\sigma_n^{-2}) = x_1\sigma_2^{-2}\sigma_3^{-2}\cdots\sigma_n^{-2} + x_2\sigma_1^{-2}\sigma_3^{-2}\cdots\sigma_n^{-2} + \cdots\; x_n\sigma_1^{-2}\sigma_2^{-2}\cdots\sigma_{n-1}^{-2}$$

$$x = \frac{x_1\sigma_2^{-2}\sigma_3^{-2}\sigma_n^{-2} + x_2\sigma_1^{-2}\sigma_3^{-2}\cdots\sigma_n^{-2}\cdots x_n\sigma_1^{-2}\sigma_2^{-2}\cdots\sigma_n^{-2}}{(\sigma_1^{-2}+\sigma_2^{-2}+\cdots\sigma_n^{-2})}$$

$$x\,(\sigma_1^{-2}+\sigma_2^{-2}+\cdots\sigma_n^{-2}) = x_1\sigma_1^{-2}+x_2\sigma_2^{-2}+\cdots\; x_n\sigma_n^{-2}.$$

$$\text{const. exp}\left(\frac{-(x-\tilde{x}_1)^2}{2\sigma^2} - \frac{(x-\tilde{x}_2)^2}{2\sigma^2}\right).$$

$$\text{const. exp}\left(-\frac{(x-a)^2}{2\sigma^2}\right)$$

$$\frac{(x-\tilde{x}_1)^2}{2\sigma_1^2}+\frac{(x-\tilde{x}_2)^2}{2\sigma_2^2} \overset{?}{=} \frac{(x-a)^2}{2\sigma^2}.$$

$$\Rightarrow \frac{x^2-2x\tilde{x}_1+\tilde{x}_1^2}{2\sigma_1^2}+\frac{x^2-2x\tilde{x}_2+\tilde{x}_2^2}{2\sigma_2^2}$$

$$\Rightarrow x^2\left(\frac{1}{2\sigma_1^2}+\frac{1}{2\sigma^2}\right)^2 - x\left(\frac{\tilde{x}_1}{\sigma_1^2}+\frac{\tilde{x}_2}{\sigma_2}\right) + \left(\frac{\tilde{x}_1^2}{\sigma_1^2}+\frac{\tilde{x}_2^2}{\sigma_2^2}\right)$$

$$ax^2+bx+c.$$

$$\left(x+\frac{b}{2}\right)^2 = x^2+bx+\frac{b^2}{4}.$$

$$x^2+bx+c = \left(x+\frac{b}{2}\right)^2 - \left(c-\frac{b}{4}\right)^2$$

$$\cancel{x^2\left(\frac{1}{2\sigma^2}\right.}$$

$$\therefore\ \frac{1}{2\sigma^2} = \frac{1}{2\sigma_1^2}+\frac{1}{2\sigma_2^2}.$$

$$\sigma^2 = \frac{1}{\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}} = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}.$$

$$\sigma^2 = \frac{\sigma_1^4}{2\sigma_1^2} = \frac{\sigma_1^2}{2}$$

$$\sigma = \frac{\sigma}{\sqrt{2}}$$

$$\frac{1}{\sigma^2} = \frac{1}{\sigma_1^2} + \cdots + \frac{1}{\sigma_n^2}.$$

$$\sigma^2 = \frac{1}{\frac{1}{\sigma_1^2} + \cdots + \frac{1}{\sigma_n^2}} = \frac{1}{\frac{n}{\sigma_1^2}} = \frac{\sigma_1^2}{n}.$$

Homework.

Start with some value x, $\sigma_1, \sigma_2 \cdots \sigma_n$ (some SD: given) $\longrightarrow$ input.

① simulate measurement errors.

repeat

$$\tilde{x}_1 = x + \text{gauss}() * \sigma_1.$$

$$\vdots$$

$$\tilde{x}_n = x + \text{gauss}() * \sigma_n$$

$$x_{new} = \frac{\tilde{x}_1 \sigma_1^{-2} + \tilde{x}_2 \sigma_2^{-2} + \cdots x_n \sigma^{-n}}{\sigma_1^{-2} + \sigma_2^{-2} + \cdots \sigma_n^{-n}}$$

$$\sqrt{\frac{1}{N}\Sigma(x_{new} - x)^2} \approx \frac{1}{\sigma_1^{-2} + \sigma_2^{-2} + \cdots \sigma_n^{-2}}.$$