VALUE ENGINEERING PART 2

- Probability Theory is a mathematical framework for uncertain phenomena.
- Long-term stability is obtained if experiments are repeated several times.

- von Mises definition of probability

→ Suppose we have an event A that occurs n_A times when executing an experiment m times

then $$P(A) = \frac{n_A}{m}$$

For stability $$\boxed{P(A) = \lim_{m \to \infty} \underbrace{\frac{n_A}{m}}_{\text{Relative frequency}}}$$

In practice, there is a trade-off for this limit.

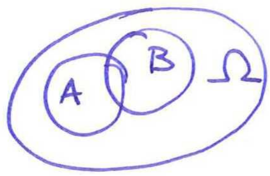
- Computational Rules

$\boxed{\Omega = \text{Full probability space: All events.}}$

3 main axioms:

1. $P(A) \geq 0$
2. $P(\Omega) = 1$
3. $P(A+B) = P(A \cup B) = P(A) + P(B)$ ← If A and B are independent

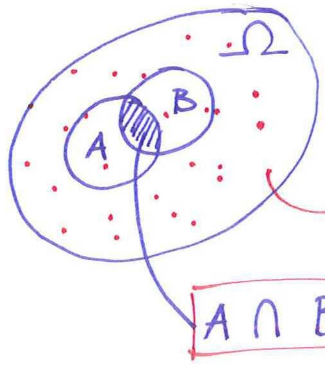
If A and B intersect $P(A \cup B) = P(A) + P(B) - P(A \cap B)$



$$P(A^c) = 1 - P(A)$$

- Impossible event ϕ → $P(\phi) = 0$

- Probability Space $\Omega = \{\omega_1, \ldots, \omega_n\}$ ω_i = outcomes of the experiments

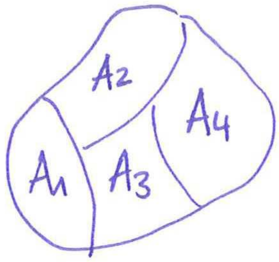


$\boxed{A \subset \Omega}$ A "is a part of" Ω

ω_i, elementary events

$\boxed{A \cap B}$

Partitions of Ω: a collection of subsets (events) A_i in such a way that the whole of Ω is covered.



$$\bigcup_{i=1}^{n} A_i = \Omega$$

All events form the total Ω.

$$A_i \cap A_j = \phi \quad (i \neq j)$$

Independent events!

$\boxed{\text{Events are subsets of } \Omega}$

(The simplest form of this? $\{A, A^c\} = \Omega$

$\boxed{\text{THEOREM: If } \Omega \text{ contains N outcomes, then there exist } 2^N \text{ possible events. (This includes } \Omega \text{ and } \phi)}$

<u>Example:</u> Throwing a coin twice

$\Omega = \{\underbrace{hh, ht, th, tt}_{2^2 \text{ events!}}\}$ $A = h$ on first toss

$P(A) = \frac{1}{2}$

<u>Homework 1</u> Exp: Throw a die twice

$A =$ Total outcome sum is 4

$P(A)$? 2 independent events

$36 = 6 \cdot 6$ possible outcomes

$P(A) = \frac{3}{36} = \frac{1}{12}$

- Countable Outcomes

Non-countable $\Omega = \{-\infty < x < +\infty\}$

Elementary events $\{x_i\}$. This # is not countable

$$P(\{x_i\}) = 0 \text{ for all } x_i$$

The probability of getting an exact value is 0.
E.g. for Ohm's law, the probability of getting exactly $V=2.3$ is 0.

- Conditional Probability

$P(A|B)$: The probability of A given that B has happened.

"given"

$$\boxed{P(A|B) = \frac{P(A \cap B)}{P(B)}}$$

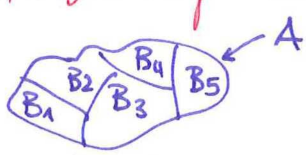
you basically normalize the intersection w.r.t. B.

- Total Probability Rule

It expresses the total probability of an outcome (A) which can be realized via several distinct outcomes (B_i)

$$\therefore \boxed{\begin{aligned} P(A) &= \sum_{i=1}^{n} P(A \cap B_i) \\ &= \sum_{i=1}^{n} P(B_i)\, P(A|B_i) \end{aligned}}$$

Using the conditional probability relation



- Bayes Rule

It makes a relationship between $P(A|B)$ and $P(B|A)$

$\rightarrow (A \cap B) = (B \cap A)$

$\therefore P(B|A)P(A) = P(A|B)P(B)$

Also a result of the conditional probability relation

$$\therefore \boxed{P(B|A) = P(A|B)\frac{P(B)}{P(A)}}$$

Example Exam with 20 Y/N questions.
A student who studies for a 6 has 0.6 probability of getting a question correct.

$A = \{\text{answer correct}\}$
$B_1 = \{\text{The student knows}\}$
$B_2 = \{\text{The student does not know}\}$

$\boxed{B_1 \text{ and } B_2 \text{ make up } A}$ $\therefore$ $P(A|B_1) = 1$ $\quad P(A|B_2) = \frac{1}{2}$

Total probability Rule $\rightarrow$ $P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2)$

$P(B_1) = 0.6 = \frac{3}{5}$ $\quad P(B_2) = P(B_1^c) = 1 - \frac{3}{5} = \frac{2}{5}$

If you study for a 6 you will get an 8 $\rightarrow$ $P(A) = 1 \cdot \frac{3}{5} + \frac{1}{2} \cdot \frac{2}{5} = \frac{4}{5}$

The probability the student knows the answer given that it's correct is 75% $\rightarrow$ $P(B_1|A) = \dfrac{P(A|B_1)\,P(B_1)}{P(A)} = \dfrac{1 \cdot \frac{3}{5}}{\frac{4}{5}} = \dfrac{3}{4}$

- **Independence**

Two events are independent if $P(A|B) = P(B|A)$

$\therefore$ $P(A|B) = P(A)$
$P(B|A) = P(B)$

We then see that $P(A \cap B) = P(A) \cdot P(B)$

RANDOM VARIABLES

• Random variable: A variable whose value is subject to variation due to chance.

A random variable has a value which is dependent upon the outcome of a random process.

Denoted with a bar below the variable → $\underline{x}$

In this case A is the event that the outcome is less than a number x. →

$$P(\underline{x} \leq x) = P(A) = F_{\underline{x}}(x) \leftarrow \text{Distribution function of } \underline{x}$$

$$\therefore \quad F_{\underline{x}}(-\infty) = 0 \qquad F_{\underline{x}}(+\infty) = 1$$

$$P(x_1 \leq \underline{x} \leq x_2) = F_{\underline{x}}(x_2) - F_{\underline{x}}(x_1)$$

• Discrete Random Variable

→ A random variable as above can have any value.
→ A discrete random variable, however, can only have specific values (such as the outcome of throwing dice)

! $0 \neq \boxed{P_{\underline{x}}(x = a)}$ Probability Mass Function.
→ a is a value out of a discrete set.

For least squares, as we are mostly dealing with measurements we will use mostly continuous random variables.

Homework 2: Toss a coin 3 times → 2^3 possible outcomes

Random variable $\underline{x}$ = # of heads

a. Domain and range of $\underline{x}$
b. PMF of $\underline{x}$
c. Cumulative Distribution function (CDF) of $\underline{x}$

a. Range $0 \leq \underline{x} \leq 3$, Domain $[0,3]$
b. PMF, Binomial: $P_{\underline{x}}(k) = \frac{n!}{(n-k)!\,k!} p^k (1-p)^{n-k}$, $p = \frac{1}{2}$ $n = 3$ $k = 0,1,2,3$
c. CDF $P_{\underline{x}}(x) = P(\underline{x} \leq 3)$

- **Continuous Random Variable**

$F_{\underline{x}}(x)$ is a continuous function, not a piecewise function, in this case.

Going back to countable outcomes, it is thus impossible to get an exact value $P(\underline{x} = x) = 0$

Probability Density Function $\boxed{F_{\underline{x}}(x) = \int_{-\infty}^{x} f_x(x')\,dx'}$

case $\underline{x} \le x$

Cumulative Density Function $\boxed{f_{\underline{x}}(x) = \frac{d}{dx} F_{\underline{x}}(x)}$

$$\therefore \quad \boxed{P(x_1 \le \underline{x} \le x_2) = \int_{x_1}^{x_2} f_{\underline{x}}(x')\,dx'}$$

- **Properties of PDF**

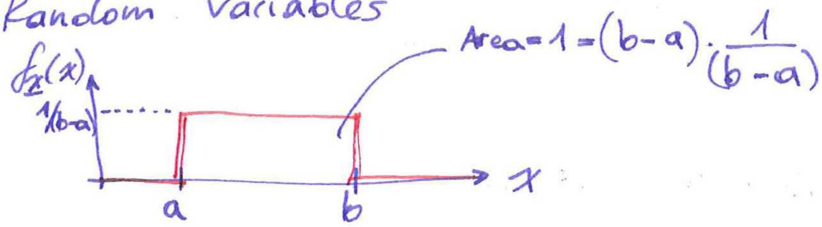
$f_{\underline{x}}(x) \ge 0$ ← There is always some probability

$\int_{-\infty}^{+\infty} f_{\underline{x}}(x) = 1$ ← Total probability for all events is 1

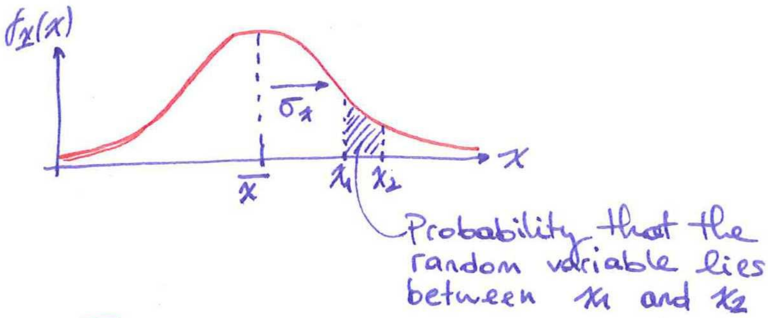
↑ These are really similar to the initial probabilities.

- Examples of continuous Random variables

Uniformely Distributed Random Variable



Gaussian/Normal Distributed Random Function.



$$\boxed{f_{\underline{x}}(x) = e^{-\frac{(x-\bar{x})^2}{2\sigma_x^2}} \cdot \frac{1}{\sigma_x\sqrt{2\pi}}}$$

The distribution is dependent on $\bar{x}$ and σ_x

There is a trick to remove the dependency on $\underline{x}$ and σ_x

↳ The Standard Normal Distribution !

$$\underline{z} = \frac{x - \bar{x}}{\sigma_x} \qquad \boxed{F_{\underline{z}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}}$$

$\underline{z}$ is a transformation random variable.

In the exam, the table gives the CDF of $\underline{z}$, denoted $\phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$

- Expected Value $\boxed{\bar{x} = E(\underline{x}) = \int_{-\infty}^{+\infty} x f_{\underline{x}}(x)\, dx}$

Expectation operator

$\boxed{\text{THEOREM}: \quad \underline{y} = g(\underline{x})}$ A function of a random variable is a random variable.

Examples $\therefore \quad \bar{y} = \int_{-\infty}^{+\infty} y f_{\underline{y}}(y)\, dy = \int_{-\infty}^{+\infty} g(x) f_{\underline{x}}(x)\, dx$

- Standard Deviation $P(|\underline{x} - \bar{x}| \leq 1.96\sigma_x) = 0.95$

- Random Vector $\vec{\underline{x}} = \begin{bmatrix} \underline{x_1} \\ \underline{x_2} \\ \vdots \\ \underline{x_n} \end{bmatrix}$ $\qquad \Omega \rightarrow \mathbb{R}^n$

Domain → Range

All probabilities must be defined in the same probability space, they must relate to the same thing.

$\therefore$ Joint CDF $P(\underline{x_1} \leq x_1, \underline{x_2} \leq x_2, \ldots, \underline{x_n} \leq x_n) = P(\underline{x_i} \leq x_i)$
$= F_{\underline{x}}(x_1, x_2, \ldots, x_n)$

Properties from the single case are retained!

- $F_{\vec{\underline{x}}}(\infty, \infty, \ldots, \infty) = 1$
- $f_{\vec{\underline{x}}} \geq 0$

It is possible to from a joint CDF to a marginal CDF (the CDF for one vector item), but not the other way around.

- Marginal CDF $\quad F_{\underline{x}_i}(x_i) = F_{\underline{x}}(\infty, \infty, \dots, x_i, \dots, \infty)$

Marginal $\quad$ Joint

This is a way of isolating the effects of only one variable (or more!)

$$F_{\underline{x}}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} \underbrace{f_{\underline{x}}(u_1, \dots, u_n)}_{\text{Joint Probability Density Function}} du_1 \dots du_n$$

$$f_{\underline{x}}(x_1, \dots, x_n) = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} F_{\underline{x}}(x_1, \dots, x_n)$$

Getting the marginal PDF out of $f(x_1, \dots, x_4)$

$$f_{\underline{x}_2}(x_2) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x_1, x_2, x_3, x_4)\, dx_1\, dx_3\, dx_4$$

Marginal $\quad$ Joint

You integrate the variables you are not interested in.

If the random variables are independent

$$f_{\underline{x}}(x_1, x_2, x_3, \dots, x_n) = f_{\underline{x}_1}(x_1) \cdot f_{\underline{x}_2}(x_2) \cdot f_{\underline{x}_3}(x_3) \cdots f_{\underline{x}_n}(x_n)$$

It is then possible to go from marginal to joint, only in this special case.

- Covariance

Covariance is a measure of dependence of 2 random variables

$$C(x_1, x_2) = E\left[(\underline{x}_1 - \overline{x}_1)(\underline{x}_2 - \overline{x}_2)\right] = (\underline{x}_1 - \overline{x}_1)(\underline{x}_2 - \overline{x}_2) f(x_1, x_2)$$

$$\therefore \quad \boxed{C(x_1, x_2) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x_1 - \overline{x}_1)(x_2 - \overline{x}_2) f(x_1, x_2)\, dx_1\, dx_2}$$

For 2 independent random variables $\quad C(x_i, x_j) = 0$

$C(x_1, x_2) = E\left[(\underline{x}_1 - \overline{x}_1)(\underline{x}_2 - \overline{x}_2)\right] = E(\underline{x}_1 \underline{x}_2) - \overline{x}_1 \overline{x}_2$

If independent $\quad f(x_1, x_2) = f(x_1) f(x_2) \longrightarrow E(\underline{x}_1 \underline{x}_2) = \overline{x}_1 \overline{x}_2$

$$C(\underline{x_1}, \underline{x_2}) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x_1 - \bar{x}_1)(x_2 - \bar{x}_2)\, f_{\underline{x}}(x_1, x_2)\, dx_1\, dx_2$$

If x_1 and x_2 are independent

$$C(\underline{x_1}, \underline{x_2}) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x_1 - \bar{x}_1)(x_2 - \bar{x}_2)\, f_{x_1}(x_1)\, f_{\underline{x_2}}(x_2)\, dx_1\, dx_2$$

$$= \int_{-\infty}^{+\infty} (x_1 - \bar{x}_1)\, f_{\underline{x_1}}(x_1)\, dx_1 \cdot \int_{-\infty}^{+\infty} (x_2 - \bar{x}_2)\, f_{x_2}(x_2)\, dx_2$$

$$= (\bar{x}_1 - \bar{x}_1) \cdot (\bar{x}_2 - \bar{x}_2) = 0$$

$$\left(\Longrightarrow \int x_1 f_{\underline{x_1}}(x_1)\, dx_1 = \bar{x}_1 \right)$$

For a set of variables $\quad \underline{u} = \underline{x_1} + \dots + \underline{x_n}$

Relation to standard deviation for the joint set. $\quad \sigma_u^2 = \sum_{i=1}^{n} \sigma_i^2 + \sum_{i<j} \sum_{j<n} C(x_i, x_j)$

Homework 3

$x_1, x_2, \dots, x_n$ are independent with variance σ^2

What is the variance for $\quad \underline{u} = \frac{1}{n} \sum_{i=1}^{n} \underline{x_i}$?

- Correlation

$$\xi(\underline{x_1}, \underline{x_2}) = \frac{C(\underline{x_1}, \underline{x_2})}{\sigma_{x_1}\sigma_{x_2}} \qquad -1 \leq \xi \leq 1$$

Correlation is a weighted covariance

- Variance - Covariance Matrix

To put things into perspective → $C(\underline{x_1}, \underline{x_1}) = \sigma_{x_1}^2$

This matrix spans the covariance of all variables

$$C = \begin{bmatrix} C(\underline{x_1}, \underline{x_1}) & C(\underline{x_1}, \underline{x_2}) & \cdots & C(\underline{x_1}, \underline{x_n}) \\ \vdots & C(\underline{x_2}, \underline{x_2}) & & \vdots \\ & & \ddots & \\ C(x_n, x_1) & \cdots & & C(\underline{x_n}, \underline{x_n}) \end{bmatrix}$$

← Symmetric Matrix

keep in mind that $C(x_i, x_j) = C(x_j, x_i)$

LEAST-SQUARES METHOD

- The objective is to estimate unknown parameters
- This is done from uncertain data
- \# data points ≥ # unknowns → Redundancy
  - ↳ Gives the possibility of determining the precision of the unknowns
  - ↳ Allows us to check for errors in our data

- We will work with inconsistent equations
  i.e. No exact solution due to uncertainty of data

The data is a set of random variables

$$\underline{y} = (\underline{y_1}, \underline{y_2}, \ldots, \underline{y_n})^T$$

We will aim to use this to fit the function: $y = A(x)$

m-dimensional vector representing y

n-dimensional vector with unknown parameters.

We will have $m \gg n$

In reality, we will never get an exact match:

$\therefore \quad y \approx A(x) \quad$ or $\quad y = Ax + \underline{e}$

where $E(\underline{e}) = 0$ if experiment is repeated several times.

A is a function that transforms $\mathbb{R}^n \to \mathbb{R}^m$
It uses an assumed function model.
In these lectures we will confine ourselves to the linear systems.

$$E(y) = A \cdot \vec{x}$$

Now that the problem is mapped out, we can proceed to look at solutions.

<u>Example</u> Device measurements for altitude

Device 1 $x_1 \approx x$
$\vdots$
Device n $x_n \approx x$

x is the true altitude but not all devices get it exactly right.

$\therefore \quad y_i = x + \underline{e_i}$

Now assume $\quad e_i \sim N(0, \sigma^2)$

Denotes a PDF

Same for all devices

$\therefore \quad y_i \sim N(x, \sigma^2) \quad \leftarrow$ If you add a number to a gaussian random variable then you simply add the mean.

For such a case, we can find that

$$\boxed{\hat{x} = \frac{1}{m} \sum_{i=1}^{m} x_i} \qquad \boxed{\sigma_{\hat{x}} = \frac{\sigma}{\sqrt{n}}}$$

<u>The average</u>: This won't be exact but it's the best we can do.

We could also redefine the problem, and state that the accuracy for each device is different.

$$e_i \sim N(0, \sigma_i^2)$$

$$y_i \sim N(x, \sigma_i^2)$$

changes for each device.

Assuming a linear model $y = A\vec{x}$ $\quad A = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$

- Redundancy $= m - n = m - 1$

$n = 1$. This is the size of x, which is 1 as there is only 1 result.

LEAST-SQUARES, NON-PROBABILISTIC, LINEAR

$$\underset{m\times 1}{y} \approx \underset{m\times n}{A}\,\underset{n\times 1}{x}$$

$m \geq n = \text{rank}(A)$

Invertible!

- If $m = n$ $\quad \hat{x} = A^{-1}y \rightarrow$ Consistent Equations
- If $m > n \longrightarrow$ Inconsistent Equations

$$y \notin col(A)$$

Redundancy $= m - n > 0 \leftarrow$ Overdetermined!

Example

$$\left.\begin{array}{l} y_1 = 2x \\ y_2 = 3x \\ y_3 = 4x \end{array}\right\} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} x$$

Redundancy $= 3 - 1 = 2$
Overdetermined!

The exact solution will require $\vec{y} = A \cdot \text{constant}$.

Eq. 1 $\rightarrow e_1 = 2x - y_1$
Eq. 2 $\rightarrow e_2 = 3x - y_2$
Eq. 3 $\rightarrow e_3 = 4x - y_3$

The errors must be minimized, to minimize the distance for all points

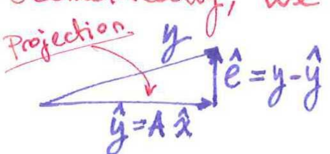
$$|Ax - y|^2 = (2x - y_1)^2 + (3x - y_2)^2 + (4x - y_3)^2 = (Ax - y)^T(Ax - y)$$

To minimize we have ~~to~~ find $\frac{\partial}{\partial x}(|Ax - y|^2) = 0$

$$\therefore \quad 0 = 2\left[(2x - y_1)2 + (3x - y_2)3 + (4x - y_3)4\right]$$

$$\Rightarrow \hat{x} = \frac{2y_1 + 3y_2 + 4y_3}{2^2 + 3^2 + 4^2} = \frac{A^T y}{|A|} = \underbrace{(A^T A)^{-1} A^T y}_{\text{Least squares solution!}}$$

Geometrically, we aim for $\hat{e} \rightarrow 0$



$\hat{e}$ and $\hat{y}$ are perpendicular

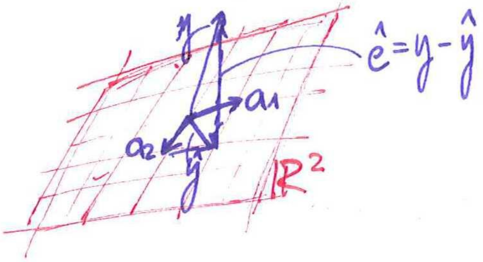
- The $\mathbb{R}^2$ case: consider $m=3 \quad n=2$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Red. $= 3-2 = 1$

Col(A) spans $\mathbb{R}^n$

We aim to find the smallest projection on the plane $\mathbb{R}^2$ that has the closest distance (i.e. perpendicular) to y.



$A^T \hat{e} = 0$ $\quad$ $\hat{e}$ is perpendicular to a_1 and a_2

$$0 = A^T \hat{e} = A^T(y - \hat{y}) = A^T(y - A\hat{x})$$

We found before that $\hat{x} = (A^TA)^{-1}A^Ty$

$$\therefore \quad 0 = A^T(y - A(A^TA)^{-1}A^Ty)$$

From $y = A\hat{x}$ it follows $\boxed{A^TA\hat{x} = A^Ty}$

Multiplying both sides of an inconsistent set with A^T eliminates the inconsistency! We now have: Red $= m-n = 0$!

We can now proceed to find an exact solution.

BUT $(A^TA)^{-1}$ $\rightarrow$ Only exists if Rank$(A) = n$
$\rightarrow$ All columns of A have to be linearly independent.

Homework 4

Solve $\quad 10 = 3x$
$\quad\quad\quad 5 = 4x$

Check the solution
Check $\hat{e} \perp \begin{bmatrix} 3 \\ 4 \end{bmatrix}$

VARIANCE PROPAGATION LAW

$\vec{y} = a\underline{\vec{x}} + b$ then $\sigma_y^2 = a^2\sigma_x^2$

We are working with random variables again!

The variance-covariance matrix of $\vec{x}$ is Q_x

$$Q_y = \int [(y-\bar{y})(y-\bar{y})^T] f_y(y)\,dy$$

→ Can also be $f_{\underline{x}}(x)\,dx$

$$= \int A(x-\bar{x})\underbrace{[A(x-\bar{x})]}_{=(x-\bar{x})^T A^T}{}^T f_{\underline{x}}(x)\,dx$$

$$= A\left[\int (x-\bar{x})(x-\bar{x})^T f_{\underline{x}}(x)\,dx\right] A^T$$

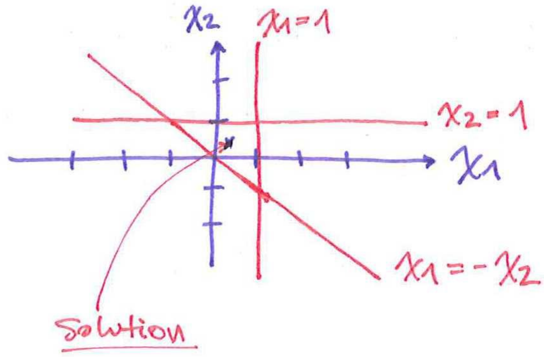
$$\Rightarrow \boxed{Q_y = A\, Q_x A^T}$$

INCONSISTENT SETS

$1 = x_1$
$1 = x_2$
$0 = x_1 + x_2$

} This set is inconsistent.

$$\underbrace{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\vec{x}}$$



$\hat{\vec{x}} = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = (A^TA)^{-1}A^Ty$, $A^TA = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$, $(A^TA)^{-1} = \frac{1}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$

$\hat{\vec{x}} = \begin{bmatrix} 1/3 \\ 1/3 \end{bmatrix}$

We found a solution to the inconsistent system! This solution is a match for all... as closely as possible.

$\hat{e} = y - \hat{y} = y - A\hat{x} = \begin{bmatrix} 2/3 \\ 2/3 \\ -2/3 \end{bmatrix}$ $e \perp A$? $e \times A = [0\ \ 0]$

The least square residual vector (e) is perpendicular to A.

Note: If $\operatorname{rank}(A) = n = m$, $(A^T)^{-1} = {A^T}^{-1}$

$\therefore\ \hat{x} = A^{-1}\cancel{(A^T)^{-1}A^T}y \longrightarrow \boxed{\hat{x} = A^{-1}y}$

Homework 5

Projection Matrix $P = A(A^TA)^{-1}A^T$

Verify: $P^2 = P$, $P^T = P$

LEAST-SQUARES LINE FITTING

Assume measurements for distance $\vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ where $y_i = y_0 + v t_i$

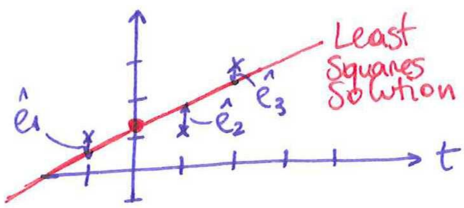
Moving at constant v.

$$\therefore \quad \vec{y} = A\vec{x} \Rightarrow \underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_n \end{bmatrix} \begin{bmatrix} y_0 \\ v \end{bmatrix}}_{\text{Model that we want to fit.}}$$

t_1 (s)	y_i (m)
-1	1
1	1
2	3

$$\Longrightarrow \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} y_0 \\ v \end{bmatrix}$$

$$\hat{x} = \begin{bmatrix} \hat{y}_0 \\ \hat{v} \end{bmatrix} = (A^TA)^{-1}A^Ty = \begin{bmatrix} 9/7 \\ 4/7 \end{bmatrix}$$



Where $e_i = y_i - Ax_i$

$$\sum_{i=1}^{m} e_i^2 = (y - Ax)^T(y - Ax) = |y - Ax|^2$$

- Using Weight: This allows us weighing some measurements as more important than others.

$$W = \begin{bmatrix} w_1 & \phi \\ & \ddots & \\ \phi & & w_m \end{bmatrix}$$

Then $\boxed{\hat{x} = (A^TWA)^{-1}A^TWy}$ if $m > n$

$\boxed{\hat{x} = A^{-1}\cancel{W^{-1}(A^T)^{-1}A^TW}y = A^{-1}y}$ if $m = n$

Then we are independent of weights!

Take $y_1 \approx x$ $\quad$ $\begin{matrix} y_1 \\ y_2 \end{matrix} \approx A\vec{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \vec{x}$ $\quad$ $r = m - n = 2 - 1 = 1$

$y_2 \approx x$ $\quad$ Inconsistency!

$\rightarrow$ without W $\quad$ $\hat{x} = (A^T A)^{-1} A^T y$, $\quad A^T A = [1\ 1] \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 2$

$(A^T A)^{-1} = 1/2$

$$\hat{x} = \frac{1}{2} [1\ 1] \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

$$\hat{x} = \frac{1}{2} y_1 + \frac{1}{2} y_2$$ $\leftarrow$ So the average

$\rightarrow$ With $W = \begin{bmatrix} w_1 & 0 \\ 0 & w_2 \end{bmatrix}$ $\quad$ $\hat{x} = (A^T W A)^{-1} A^T W y$, $\quad A^T W A = [1\ 1] \begin{bmatrix} w_1 & 0 \\ 0 & w_2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = w_1 + w_2$

$(A^T W A)^{-1} = \frac{1}{w_1 + w_2}$

$$\therefore \hat{x} = \frac{1}{w_1 + w_2} \begin{bmatrix} w_1 & 0 \\ 0 & w_2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

$$\hat{x} = \frac{w_1 y_1 + w_2 y_2}{w_1 + w_2}$$ $\leftarrow$ So the weighted average.

Homework 6

$y = \begin{bmatrix} 3.5 \\ 0.5 \\ 3.7 \\ 2.8 \end{bmatrix}$ $\quad A = \begin{bmatrix} 1 & 2 \\ 1 & 0 \\ 1 & 3 \\ 1 & 1 \end{bmatrix}$ $\quad y = y_0 + vt$

$W = 3 I_4$ All weights are the same so they might as well not be there!

• Base Functions

Take $y_i \approx f(t_i)$ $\quad i = 1, \ldots, m$

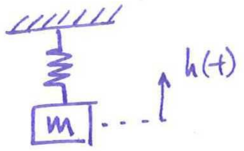
Function f can be written as a linear combination of n known base functions

$$\therefore f(t) = \sum_{a=1}^{m} x_a \phi_a(t) = x_1 \phi_1(t) + x_2 \phi_2(t) + \ldots + x_n \phi_n(t)$$

Or, in matrix form $\rightarrow$ $\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \underbrace{\begin{bmatrix} \phi_1(t_1) & \cdots & \phi_n(t_1) \\ \vdots & \ddots & \vdots \\ \phi_1(t_m) & \cdots & \phi_n(t_m) \end{bmatrix}}_{m \times n} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$

Homework 8

* There is no Homework 7 ...



$$\omega = \sqrt{\frac{k}{m}}$$

$$\omega = 10 \ \frac{\text{rad}}{s}$$

$$h(t) = a \underbrace{\cos wt}_{\phi_1(t)} + b \underbrace{\sin wt}_{\phi_2(t)}$$

t(s)	-0.4	-0.2	0	0.2	0.4
h(m)	-3	-16	6	9	-8

$$\begin{bmatrix} h_1 \\ \vdots \\ h_5 \end{bmatrix} \approx \underbrace{\begin{bmatrix} \cos wt_1 & \cdots & \sin wt_1 \\ \vdots & \ddots & \vdots \\ \cos wt_5 & \cdots & \sin wt_5 \end{bmatrix}}_{\substack{5 \times 2 \\ (n \times m)}} \begin{bmatrix} a \\ b \end{bmatrix}$$

Homework 9

y	1.5	3.5	6.2	3.2
t_1	5	3	5	3
t_2	0.5	0.5	0.3	0.3

Model 1: $y = a_0 + a_1 t_1 + a_2 t_2$

Model 2: $y = a_0 + a_1 t_1 + a_2 t_2 + a_3 t_1 t_2$

Random Errors

$\vec{y} = A\vec{x} + \vec{e}$ → Error random vector

This is not to be confused with $\hat{e}$

$\hat{e}$ indicates an error with respect to a unique solution

Measurement errors are modelled with y and $\underline{e}$

$E(\underline{e}) = 0$

$\therefore E(\underline{y}) = Ax$

It is thus assumed that, on average, measurement errors are such that you can extract the true model.

$$\underline{y} = Ax + \underline{e} \rightarrow \hat{\underline{x}} = (A^T W A)^{-1} A^T W (Ax + \underline{e})$$

$$= (A^T W A)^{-1} [A^T W A x + A^T W \underline{e}]$$

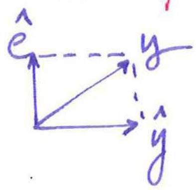
$$\boxed{\hat{\underline{x}} = x + (A^T W A)^{-1} A^T W \underline{e}}$$ Unbiased Estimator.

Then it also follows that $E(\hat{\underline{x}}) = x$

An unbiased estimator aims to use weights so as to minimize the error term and get as close as possible to the the real value.

$$\hat{\underline{y}} = A\hat{\underline{x}} = Ax + \underbrace{A(A^T W A)^{-1} A^T W}_{= P \text{ (Projection matrix)}} \underline{e}$$

Once again the solution is written as a deterministic part (Ax) and a random error part ($P\underline{e}$)



$$\boxed{\hat{e} = \underline{y} - \hat{\underline{y}} = 0 + (I_m - P)\underline{e}}$$

Determining the Accuracy

Assumptions : You know everything about the measurements

: You even know the variances and covariances of the vectors, at least assume $\underline{Q_y}$ is known

If you know Q_y you can find Q_x

$\hat{\underline{x}} = By$, $B = (A^T W A)^{-1} A^T W$ $\quad Q_{\hat{x}} = B Q_{y} B^T$

$$\therefore Q_{\hat{x}} = [(A^T W A)^{-1} A^T W] \, Q_y \, [(A^T W A)^{-1} A^T W]^T$$

So if Q_y is known we can get $Q_{\hat{x}}$ before we even take the measurements. This will inform us on the precision of the parameters and, if necessary, indicate that the experiment should be revised.

Example

$$y_1 \approx x, \quad y_2 \approx x, \quad W = \begin{bmatrix} w_1 & 0 \\ 0 & w_2 \end{bmatrix} \rightarrow \hat{x} = \frac{w_1 y_1 + w_2 y_2}{w_1 + w_2}$$

$$A = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad Q_y = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} = \sigma^2 I_2$$

What will be the value of $\sigma_{\hat{x}}^2$?

$$A^T W A = [1\ 1] \begin{bmatrix} w_1 & 0 \\ 0 & w_2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = w_1 + w_2 \qquad \text{(Scalar)}$$

$$\sigma_{\hat{x}}^2 = Q_{\hat{x}} = \frac{1}{w_1+w_2} [1\ 1] \begin{bmatrix} w_1 & 0 \\ 0 & w_2 \end{bmatrix} \sigma^2 \begin{bmatrix} w_1 & 0 \\ 0 & w_2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \frac{1}{w_1+w_2}$$

$$\sigma_{\hat{x}}^2 = \sigma^2 \underbrace{\frac{w_1^2 + w_2^2}{(w_1+w_2)^2}}$$

This is how the variance of the measurements propagates to the variance of the solution.

$$\text{If } w_1 = w_2 \longrightarrow \sigma_{\hat{x}}^2 = \frac{\sigma^2}{2}$$ } Best possible option, mathematically.

$$\text{If } w_1 \gg w_2 \longrightarrow \sigma_{\hat{x}} = \sigma^2$$

Given $Q_y = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$ What is W to minimize variance of $\hat{x}$?

$\boxed{W = Q_y^{-1}}$ This is the best we can do, mathematically

<u>THEOREM</u> : Gauss - Markov Theorem

If $E(\underline{y}) = Ax$ and Q_y given

$$\hat{x} = (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} \underline{y}$$

$$\therefore W = Q_y^{-1}$$

Best Linear Unbiased Estimator

BLUE

Then:
$$Q_{\hat{x}} = (A^T Q_y^{-1} A)^{-1} \cancel{A^T Q_y^{-1} Q_y Q_y^{-1} A (A^T Q_y^{-1} A)^{-1}}$$

$$\boxed{Q_{\hat{x}} = (A^T Q_y^{-1} A)^{-1}}$$

If $Q_y = \sigma^2 I_m \longrightarrow \boxed{Q_{\hat{x}} = \sigma^2 (A^T A)^{-1}}$ Even simpler

• Defining $Q_{\hat{y}}$

$$\hat{y} = A\hat{x} \rightarrow Q_{\hat{y}} = AQ_{\hat{x}}A^T = A(A^TQ_y^{-1}A)^{-1}A^T$$

$$\underline{\hat{e}} = y - \hat{y} \rightarrow \boxed{Q_{\hat{e}} = Q_y - Q_{\hat{y}}}$$

$$\boxed{Q_e = Q_y}$$

This is only true for BLUE (ie $W = Q_y^{-1}$)
Needed for blunder detection, definitely on exam.

MULTIVARIATE NORMAL DISTRIBUTION

$$f_{\underline{x}}(x) = \frac{1}{\sigma_x\sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma_x^2}}$$

$$\therefore \text{ JOINT PDF: } f_{\underline{x}}(x) = \frac{1}{\sqrt{|2\pi Q_x|}} e^{-\frac{1}{2}\underbrace{(x-\bar{x})^T}_{1\times n}\underbrace{Q_x^{-1}}_{n\times n}\underbrace{(x-\bar{x})}_{n\times 1}}$$

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

Where $Q_x = \begin{bmatrix} \sigma_1^2 & & \emptyset \\ & \ddots & \\ \emptyset & & \sigma_n^2 \end{bmatrix}$

$x_1, \dots, x_n$ are independent. There are no covariances!

$$(x-\bar{x})^T Q_x^{-1}(x-\bar{x}) = \sum_{i=1}^{n} \frac{(x_i - \bar{x}_i)^2}{\sigma_i^2}$$

← Makes sense :)

$$|2\pi Q_x| = 2\pi^n \prod_{i=1}^{n} \sigma_i^2$$

Then we get $f_{\underline{x}}(x) = \prod_{i=1}^{n} \left[\frac{1}{\sqrt{2\pi}\,\sigma_i} e^{\frac{-(x_i-\bar{x})^2}{2\sigma_i^2}} \right]$

Chi-Squared (χ^2) Distribution

- The distribution of the sum of squares of m independent normal random variables.
- Used to determine the "goodness of fit" of an observed distribution to a theoretical one.

Assumptions: $\underline{x} \sim N(0, Q_x)$

$$Q_x = \sigma^2 I_m$$

$$f_{\chi^2}(t) = \begin{cases} \dfrac{x^{\frac{m}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{m}{2}} \Gamma\left(\frac{m}{2}\right)} & t \geq 0 \\ 0 & t < 0 \end{cases}$$

Mean: $\overline{\chi}^2 = k$

Variance: $\sigma^2_{\chi^2} = 2k$

- **Error Distribution**

$$\hat{e}^T Q_x^{-1} \hat{e} \sim \chi^2(\underbrace{m-n}_{=\text{redundancy}})$$

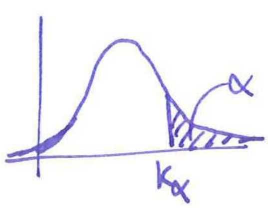
$$\therefore \quad \hat{e}^T Q_x^{-1} \hat{e} = \frac{1}{\sigma^2} \sum_{i=1}^{m} e_i^2 = \sum_{i=1}^{m} \left(\frac{\hat{e}_i}{\sigma}\right)^2 \sim \chi^2(m-n)$$

- **Significance Level**

We establish a significance level for blunder detection.

Sign. level $\alpha = 0.01$ (1%)

One-sided test

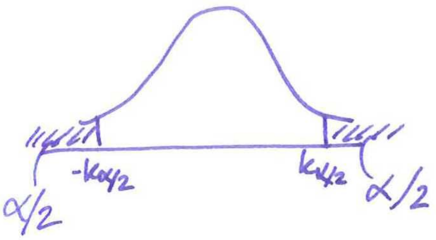


If $\boxed{\sum_{i=1}^{m} \frac{\hat{e}_i^2}{\sigma^2} > k_\alpha}$

then we must reject the model.

A smaller α means being more thorough with the model.

• 2-sided Blunder detection



→ If we have made up our mind about a model, then we need to select outliers.

→ We can say that anything that lies outside of the $\left[-\frac{k_\alpha}{2} \quad \frac{k_\alpha}{2}\right]$ interval is an outlier

$$\therefore \left|\frac{e_i}{\sigma_{e_i}}\right| > k_\alpha \rightarrow \text{Reject measurement.}$$

Squares of this are to be found on the diagonal of $Q_{\hat{e}}$