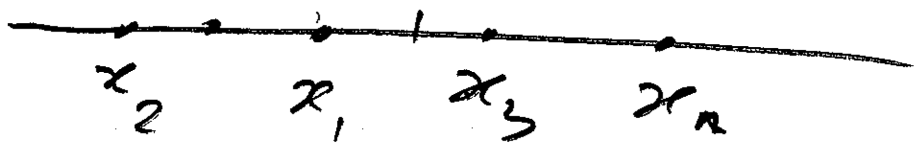
#2

$x_1, x_2 \dots x_n$ - data

<u>Center of $\{x_i\}$</u> $\qquad x$ - center



(a) $$\sum_{i=1}^{n} |x - x_i| \sim \min_x$$

$$x = \tilde{x} = \underline{\text{median}}(x_1, \dots x_n)$$

(b) $$\sum_{i=1}^{n} (x - x_i)^2 \sim \min_x$$

$$x = \bar{x} = \underline{\text{mean}}(x_1, \dots, x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

(c) $$\max_{i=\overline{1,n}} |x - x_i| \sim \min_x$$

$$x = \hat{x} = \underline{\text{middle}}(x_1, \dots x_n) =$$
$$= \text{center of extremes}$$
$$= \frac{1}{2}\left(\min_i x_i + \max_i x_i\right)$$

a) median

$x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5$

$\underbrace{\cdot \quad \cdot \quad \cdot}_{\tilde{x}_3} \quad \cdot \quad \cdot$

need to keep all of them.

Car. Info: $ord(x_1 .. x_n)$ – ordered seq.

1. Very easy to use
2. Updating: adding x_{n+1} to a sorted list – easy
3. Combining – rel. easy.

a') $x_i \in \{z_1, z_2 .. z_m\}$ – finite set of values.

$c_k = (\text{number ox } x_i = z_k) =$
$= |\{i : x_i = z_k\}|$

$|\cdot|$ – number of el-ts in a set.

Update

$\left.\begin{matrix} x_1 \\ \vdots \\ x_n \end{matrix}\right] \to (c_1, \ldots c_m)$ vector of counts.

$x_{n+1} \to \oplus \to (c_1, \ldots c_k + 1, .. c_m)$

where k: $x_{n+1} = z_k$

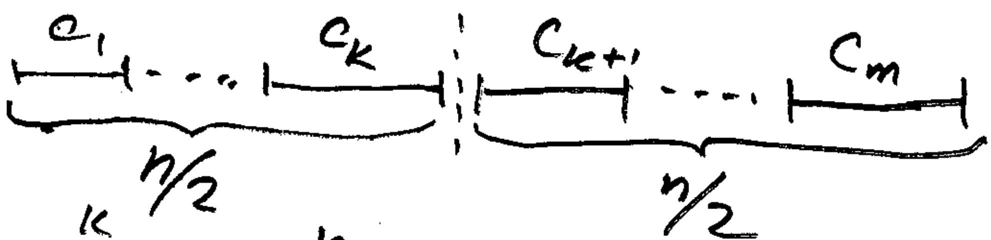
$$\left.\begin{matrix}x_1\\ \vdots \\ x_n\end{matrix}\right] \to (C_1, \dots, C_m)$$

$$\left.\begin{matrix}y_1\\ \vdots \\ y_k\end{matrix}\right] \to (D_1, \dots, D_m)$$

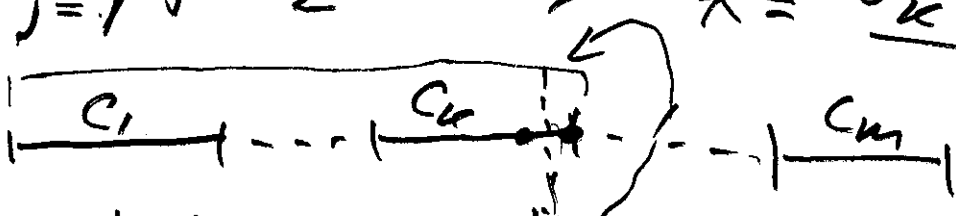
$$(C_1,\dots,C_m) \oplus (D_1,\dots,D_m) \to (C_1+D_1, \dots, C_m+D_m)$$

<u>Use</u> cau info

* n - even

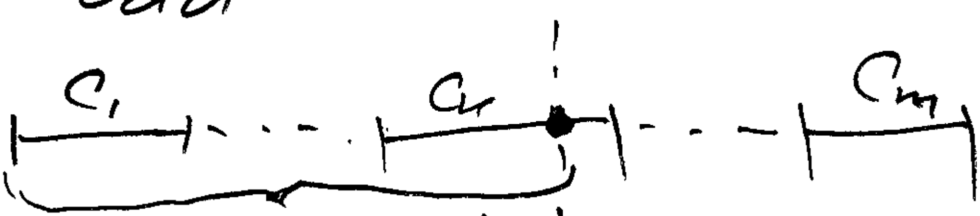


Find k: $\sum_{j=1}^{k} c_j = \frac{n}{2} \quad \Rightarrow \quad \tilde{x} = \frac{v_k + v_{k+1}}{2}$



Find k: $\sum_{j=1}^{k-1} C_j < \frac{n}{2} < \sum_{j=1}^{k} C_j \quad \Rightarrow \quad \tilde{x} = v_k$

* n - odd



Find k: $\sum_{j=1}^{k-1} C_j < \frac{n+1}{2} \leq \sum_{j=1}^{k} C_j \quad \Rightarrow \quad \tilde{x} = v_k$

b) <u>mean</u> $\bar{x}$:

can info: (n, S)

$$n = \sum_{i=1}^{n} x_i^0 \qquad S = \sum_{i=1}^{n} x_i$$

c) <u>middle</u> $\hat{x}$ (m, M)

$$m = \min_i x_i \qquad M = \max_i x_i$$

$$\hat{x} = \frac{m + M}{2}.$$

$$\left.\begin{matrix} x_1 \\ \vdots \\ x_n \end{matrix}\right] \to (m, M)$$

$$(m, M) \oplus x_{n+1} \to \left(\min(x_{n+1}, m), \max(x_{n+1}, M)\right)$$

$$\left.\begin{matrix} x_1 \\ \vdots \\ x_n \end{matrix}\right] \to (m, M) \qquad \left.\begin{matrix} y_1 \\ \vdots \\ y_k \end{matrix}\right] \to (m', M')$$

$$(m, M) \oplus (m', M') \to \left(\min(m, m'), \max(M, M')\right)$$

$(c_1 \cdots c_m)$ – can info for median.

can extract:

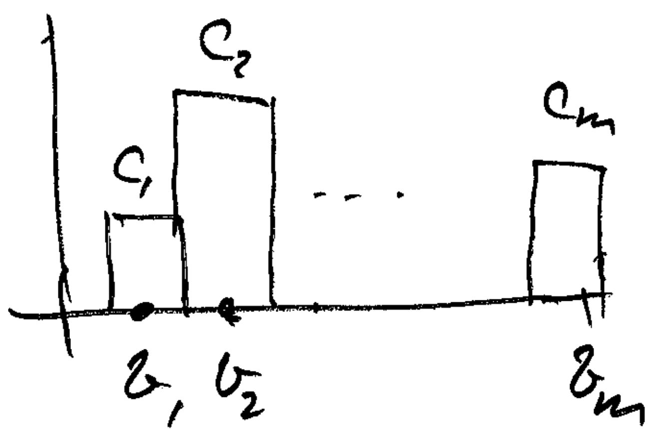
<u>mean</u>: $\bar{x} = \dfrac{\sum_{k=1}^{m} c_k v_k}{\sum_{k=1}^{m} c_k}$

<u>middle</u>:

$\hat{x} = \frac{1}{2}\left[\min\{v_k : c_k > 0\} + \max\{v_k : c_k > 0\}\right]$

<u>Histogram</u>

- common visual representation of such information.



$\Rightarrow$ $(c_1 \cdots c_m)$ can be used as can info for

$\tilde{x}$ – median
$\bar{x}$ – mean
$\hat{x}$ – middle
. . . . . .

$\Big\} = \big((c_1 \cdots c_m), (n, s), (m, M)\big)$

Canonical Information properties.

* Existence and Uniqueness.
  * Elementary info – represents 1 reading
    $x \mapsto elem(x)$
  * Empty info: repr 0 readings. (no observations)
* Combination (or composition) operation. $\oplus$

Monoid:
  * commutative: $a \oplus b = b \oplus a$
  * associative:
    $(a \oplus b) \oplus c = (a \oplus (b \oplus c))$
  * neutral element 0:
    $a \oplus 0 = a$

    can info for NO data
* Update
  $$a \oplus x = a \oplus elem(x)$$
* Compactness (minimality)
* Effectiveness
  * combination
  * updating
  * deployment.

Comment on uniqueness.

$x_1 \ldots x_n$

$(x_1\ x_2\ x_3)$ or (x_3, x_2, x_1)

* <u>Completeness</u> (Sufficiency).

Can info. should retain all the ifo which was present in the original set.

Can inf should provide the same result as the original data.

* <u>Deployment:</u>

Can Iform. $\mapsto$ final result

original algorithm

Raw Data

b) mean $\bar{X}$ for $(x_1 \ldots x_n)$

$\underline{(n, S)}$ $\quad S = \sum_{i=1}^{n} x_i$

* <u>Ex & Un.</u> <u>ok</u>
  * Elem Info $x \mapsto (1, x)$
  * Empty set $\emptyset \mapsto (0, 0)$

* <u>Combination</u>

$$(n_1, S_1) \oplus (n_2, S_2) = (n_1 + n_2, S_1 + S_2)$$

* <u>Updating</u>

$$(n, S) \oplus x = (n+1, S+x)$$

$$\Downarrow$$

$$(n, S) \oplus (1, x) = \text{---}''\text{---}$$

* Com, Assoc, Neutr. el. – trivial.

* <u>Completeness</u>

$$(n, S) \mapsto \bar{X} = \frac{S}{n}$$

works only id $n \geq 1$

* <u>compactness</u> ✓
* Efficiency
  * combining ✓
  * depl. ✓

c) <u>Middle.</u> $\hat{X} = \frac{1}{2}\left(\min_i x_i + \max_i x_i\right)$

* <u>Ex & Un.</u> (m, M) $m = \min_i X_i$

  $M = \max_i x_i$

  - Elem. Info.

    $x \longmapsto (x, x)$

  - Empty Set

    $\varnothing \longmapsto (+\infty, -\infty)$

* <u>Combination</u>

  $(m_1, M_1) \oplus (m_2, M_2) = \left(\min(m_1, m_2), \max(M_1, M_2)\right)$

  * comm ✓ * assoc ✓
  * Neutral element

  $(m, M) \oplus (+\infty, -\infty) =$

  $= \left(\underbrace{\min(m, +\infty)}_{m}, \max(M, -\infty)\right) =$

  $= (m, M).$ - ok.

* <u>Updating</u>

  $(m, M) \oplus x = \left(\min(m, x), \max(M, x)\right)$

* <u>Completeness.</u> $\hat{X} = \frac{1}{2}(m + M)$. min # of readings is $\underline{\underline{1}}$

* Compact * Efficient

a) Median. $\hat{x}$

* Ex & Un. $ord(x_1 \cdots x_n)$
* * Elem info $x \mapsto (x)$
  * Empty set $\emptyset \mapsto ()$
* Combination. $\oplus$
  merging ordered lists
  * commut v
  * assoc. v
  * Neutral el-t v
* completeness

$$\hat{x} = \begin{cases} y_{\frac{n+1}{2}} & n\text{-odd} \\ \frac{1}{2}\left(y_{\frac{n}{2}} + y_{\frac{n}{2}+1}\right) & n\text{-even} \end{cases}$$

* Compact NO! (and expanding).
* Efficiency
  * Combining not very eff.
  * deployment Ok
  * Updating rather simple.

a'). Median $\bar{x}$ when $x_i \in \{z_1 \cdots z_m\}$.

* Ex & Un.

$(x_1 \cdots x_n) \mapsto (C_1 \cdots C_m)$

* Elem. $x \mapsto (0 \cdots 1, \cdots 0)$

$k : x = z_k$

* Empty: $\emptyset \mapsto (0, \cdots 0)$

* Combine

$(C_1 \cdots C_m) \oplus (D_1 \cdots D_m) = (C_1 + D_1, \cdots C_m + D_m)$

* Completeness.

$(C_1 \cdots C_m) \mapsto \bar{x}$ need to have $n \geq 1$.

* Compactness: does not grow!

depends on goals

- relatively compact

* Efficiency. OK → combining

- deployment OK..

Info in Explicit Form

Mean $\bar{x}$ $\quad (n, \bar{x})$

* Ex & un

  Elem: $x \mapsto (1, x)$
  Empty: $\emptyset \mapsto (0, ?)$ cand: $(0, A)$
  A any => not unique.

* Combination

$$(n_1, \bar{x}_1) \oplus (n_2, \bar{x}_2) = \left(n_1 + n_2, \frac{1}{n_1 + n_2}\left[n_1 \bar{x}_1 + n_2 \bar{x}_2\right]\right)$$

  * comm ✓
  * assoc. ✓
  * Neutral element.

$$(n, \bar{x}) \oplus (0, A) = \left(n, \frac{1}{n+0}\left[n\bar{x} + 0A\right]\right) =$$

$$= (n, \bar{x}) \text{ ok. } A\text{-any.}$$

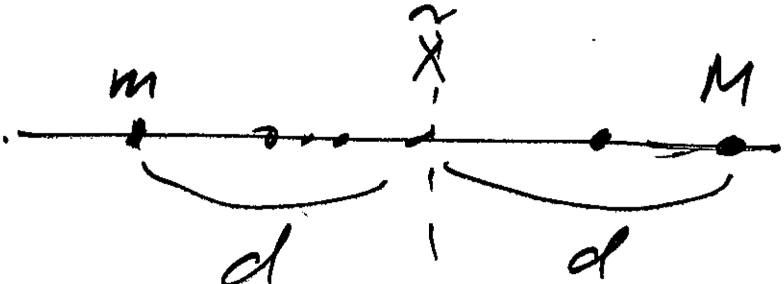
* Completeness ok by def.
* compactness ok
* Efficiency
  * combination: not bad (but not as good as it could be).
  * deployment. Very good. (the best).

Middle $(\tilde{x}, d)$

$$\tilde{x} = \frac{\max_i x_i + \min_i x_i}{2}$$

* Ex & Un.

$$d = \frac{\max x_i - \min x_i}{2}.$$



* Elem $x \mapsto (x, d)$

* Empty $\emptyset \mapsto$ (not def, not def).

* Comb.

$$(\tilde{x}_1, d_1) \oplus (\tilde{x}_2, d_2) = (\tilde{x}, d)$$

$$\tilde{x} = \frac{\max(\tilde{x}_1 + d_1, \tilde{x}_2 + d_2) + \min(\tilde{x}_1 - d_1, \tilde{x}_2 - d_2)}{2}$$

$$d = \frac{\text{—''—} \quad - \quad \text{—''—}}{2}$$

* comm
* ass
* Neutral $\emptyset = (A, -\infty)$ A - any value. should formally work

Median $\hat{x}$. $(\hat{x}$, something else$)$ ~ can info.

meaningless.

Random variables.

ν – rand var.

$\underline{\underline{E}}\nu$ – mean value

E Expectation operation.

Linear!

$$E(\alpha \nu + \beta \mu) = \qquad \alpha, \beta \text{ constants.}$$

rand vars

$$= \alpha E\nu + \beta E\mu.$$

$$\mathrm{Var}\,\nu = E(\nu - E\nu)^2 =$$

$$= E\left(\nu^2 + (E\nu)^2 - 2\nu E\nu\right) =$$

$$= E\nu^2 + (E\nu)^2 - 2 E\nu \cdot E\nu$$

$$= E\nu^2 - (E\nu)^2$$

$\nu_1, \nu_2 \dots \nu_m$ – seq of rand var.

Sample mean of ν_i

$$\bar{\nu} = \frac{1}{m} \sum_{i=1}^{m} \nu_i$$

$\{\nu_i\}$ independent, ident. distr. (i.i.d)

$E\nu_i = \mu$ – all the same

$\mathrm{Var}\, \nu_i = \sigma^2$

$\bar{\nu}$ is a "good" e. r of μ.

* <u>unbiased</u>. : $E\bar{\nu} = \mu$

$$E\bar{\nu} = E\frac{1}{m}\sum_{i=1}^{m} \nu_i = \frac{1}{m}\sum_{i=1}^{m} \underbrace{E\nu_i}_{=\mu} =$$

$$= \frac{1}{m} m\mu = \mu \quad \checkmark.$$

How can we estimate variance σ^2?

candidate: $\displaystyle \frac{1}{m}\sum_{i=1}^{m} (\nu_i - \bar{\nu})^2$