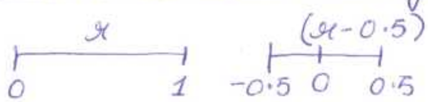
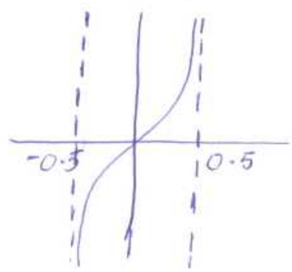
2 questions:

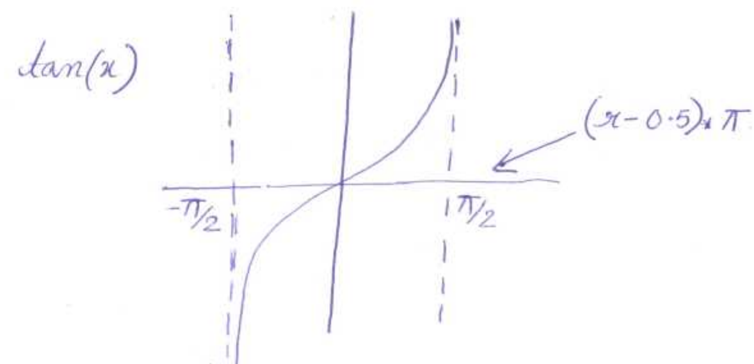
1) how to simulate Cauchy Distribution?

2) how to estimate Δ, based on sample.

How to Simulate Cachy Distribution:







c = tan((rand() - 0.5) * (Math.PI ~~/2.0~~)).

$\rho(x) \cdot \Delta x$

Cauchy Distribution with delta:

c = Delta * tan((rand() - 0.5) * (Math.PI))

Cauchy distribution w/ parameter Δ;

$$\rho(x) = \frac{1}{\pi\Delta} \frac{1}{1+\left(\frac{x}{\Delta}\right)^2}$$

1. check $\displaystyle\int_{-\infty}^{+\infty} \rho(x) \cdot dx = 1.$

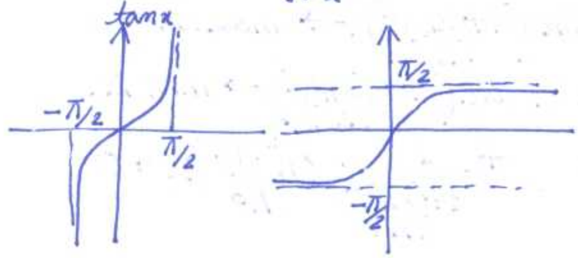
$$\int_{-\infty}^{+\infty} \frac{1}{\pi\Delta} \frac{dx}{1+\left(\frac{x}{\Delta}\right)^2} =$$

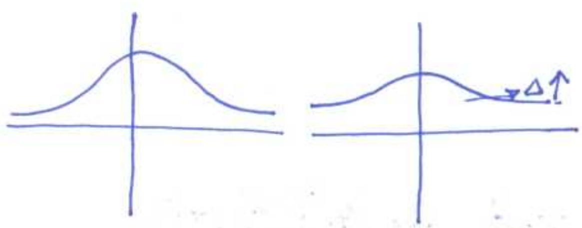
Lets simplify; $y = \frac{x}{\Delta}$, $dy = \frac{dx}{\Delta}$

$$\therefore \int_{-\infty}^{+\infty} \frac{1}{\pi} \frac{dy}{1+y^2} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{dy}{1+y^2} = \frac{1}{\pi} \cdot \text{arc.tan}(x) \Big|_{-\infty}^{+\infty} = \frac{1}{\pi} \text{arc.tan}(x) \Big]_{-\infty}^{+\infty}$$

$$\left[\because \text{arc tan } x' = \frac{1}{1+x^2} \right]$$

$$= \frac{1}{\pi} \left(\frac{\pi}{2} - \left(-\frac{\pi}{2} \right) \right) = 1.$$





$y = \tan(x)$.

$x = \text{arc.tan}(y)$. we want;

$$\frac{dx}{dy} = f(y) = \frac{1}{\frac{dy}{dx}} = \frac{1}{1/\cos^2 x} = \cos^2 x.$$

$$\frac{dy}{dx} = \frac{1}{\cos^2 x} = \frac{1}{(1+y^2)}$$

Formula:

$$y = \tan(x) = \frac{\sin x}{\cos x}, \quad \left(\frac{u}{v}\right)' = \frac{u'v - uv'}{v^2}$$

$$\therefore y' = \frac{\cos x . \cos x - \sin x . (-\sin x)}{\cos^2 x} = \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = \frac{1}{\cos^2 x}.$$

$$y^2 = \frac{\sin^2 x}{\cos^2 x} = \frac{1 - \cos^2 x}{\cos^2 x} = \frac{1}{\cos^2 x} - 1.$$

$$\therefore \frac{1}{\cos^2 x} = y^2 + 1.$$

$$\therefore \cos^2 x = \frac{1}{(1+y^2)}.$$

Maximum Likelihood:

we have $x_1, \dots x_n$.

Find: Δ for which prob. is largest.

$$\frac{1}{\pi\Delta} \cdot \frac{1}{1+\left(\frac{x_1}{\Delta}\right)^2} \cdot \frac{1}{\pi\Delta} \cdot \frac{1}{1+\left(\frac{x_2}{\Delta}\right)^2} \cdot \frac{1}{\pi\Delta} \cdot \frac{1}{1+\left(\frac{x_3}{\Delta}\right)^2} * \cdots \frac{1}{\pi\Delta} \cdot \frac{1}{1+\left(\frac{x_n}{\Delta}\right)^2} \rightarrow \max \Delta.$$

Taking the logarithm; $\qquad \ln\left(\frac{1}{b}\right) = \ln 1 - \ln b = 0 - \ln b$
$= -\ln b.$

$\frac{1}{\pi^n \cdot \Delta^n}$.

$-n \ln \Delta - \ln.\left(1+\left(\frac{x_1}{\Delta}\right)^2\right) - \ldots \quad \ln\left(1+\left(\frac{x_n}{\Delta}\right)^2\right) \to \max.\ \Delta$

$n.\ln.\Delta + \ln\left(1+\left(\frac{x_1}{\Delta}\right)^2\right) + \ldots + \ln.\left(1+\left(\frac{x_n}{\Delta}\right)^2\right) \to \min \Delta.$

$n.\frac{1}{\Delta} + \frac{1}{1+\left(\frac{x_1}{\Delta}\right)^2} \times \frac{x_1^2 \cdot (-2)}{\Delta^3} + \ldots + \frac{1}{1+\left(\frac{x_n}{\Delta}\right)^2} \cdot \frac{x_n^2 \cdot (-2)}{\Delta^3} = 0$

$\left[1 + \frac{x_1^2}{\Delta^2} = 1 + x_1^2 \cdot \Delta^{-2}\right]$

Multiplying by Δ;

$n + \frac{1}{1+\left(\frac{x_1}{\Delta}\right)^2} \times \frac{x_1^2 \cdot (-2)}{\Delta^2} + \ldots + \frac{1}{1+\left(\frac{x_n}{\Delta}\right)^2} \times \frac{x_n^2(-2)}{\Delta^2} = 0$

Simplifying;

$\frac{2x_1^2}{\left(1+\left(\frac{x_1}{\Delta}\right)^2\right)\Delta^2} + \frac{2x_2^2}{\left(1+\left(\frac{x_2}{\Delta}\right)^2\right)\Delta^2} + \ldots + \frac{2x_n^2}{\left(1+\left(\frac{x_n}{\Delta}\right)^2\right)\Delta^2} = n$

$\frac{\left(\frac{x_1}{\Delta}\right)^2}{1+\left(\frac{x_1}{\Delta}\right)^2} + \ldots + \frac{\left(\frac{x_n}{\Delta}\right)^2}{1+\left(\frac{x_n}{\Delta}\right)^2} = \frac{n}{2}.$

$\left[\text{We have} \left(\frac{z}{1+z}\right) = 1 - \frac{1}{(1+z)}\right]$

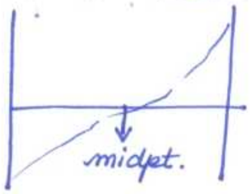
$\therefore 1 - \frac{1}{1+\left(\frac{x_1}{\Delta}\right)^2} + 1 - \frac{1}{1+\left(\frac{x_2}{\Delta}\right)^2} \ldots + 1 - \frac{1}{1+\left(\frac{x_n}{\Delta}\right)^2} = \frac{n}{2}.$

$= \frac{1}{\left(1+\left(\frac{x_1}{\Delta}\right)^2\right)} + \ldots + \frac{1}{\left(1+\left(\frac{x_n}{\Delta}\right)^2\right)} = \frac{n}{2}.$

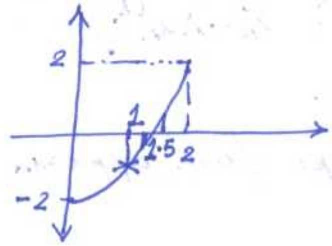
Using concept of monotonicity;

$\Delta \uparrow \quad \left(\frac{x_i}{\Delta}\right)^2 \downarrow \quad 1+\left(\frac{x_i}{\Delta}\right)^2 \downarrow \quad \frac{1}{1+\left(\frac{x_i}{\Delta}\right)^2} \uparrow$

$$\therefore F(\Delta) \stackrel{def}{=} \frac{1}{1+\left(\frac{x_1}{\Delta}\right)^2} + \cdots + \frac{1}{1+\left(\frac{x_n}{\Delta}\right)^2} = \frac{n}{2}$$

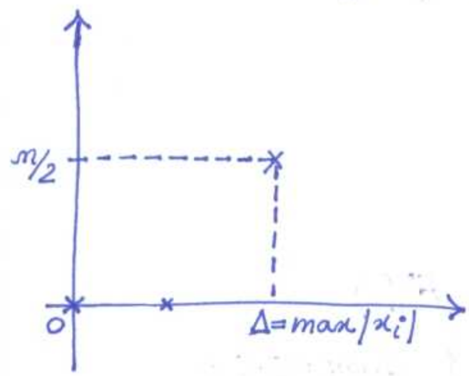


eg: $x^2 - 2 = 0$



$x \in [0, 2]$
$x \in [1, 2]$
$x \in [1, 1.5]$
$x \in [1.25, 1.5]$

When $\Delta = 0$;



we have n terms; if each is $\geq \frac{1}{2}$

$1+\left(\frac{x_i}{\Delta}\right)^2 \leq 2. \quad \left(\frac{x_i^2}{\Delta}\right) \leq 1.$

<u>Cauchy Based Monte Carlo Method</u>: (1st Approximation)

<u>Given</u>: $\tilde{x}_1, \tilde{x}_2 \cdots \tilde{x}_n, \tilde{\Delta}_1, \Delta_2 \cdots \Delta_n, f(x_1, x_2 \cdots x_n)$

<u>How</u>: we iterate N times; $k = 1, 2, \ldots N$, * for each i, we simulate Δx_i as Cauchy distribution $\Delta x_i^{(k)} = \Delta_i \cdot \tan(\pi(r - 0.5))$ $r \to$ random $\in [0,1]$

* $y^{(k)} = f(\tilde{x}_1 - \Delta x_1^{(k)}, \cdots \tilde{x}_n - \Delta x_n^{(k)})$
* $\Delta y^{(k)} = \tilde{y} - y^{(k)}$

We find Δ from $\sum_{i=1}^{N} \frac{1}{1+\left(\frac{\Delta y^{(k)}}{\Delta}\right)^2} = \frac{N}{2}$, $\Delta \in [0, \max|\Delta y^{(k)}|]$

binary search:

Correct Algorithm:

compute $\tilde{y} = f(\tilde{x}_1, \tilde{x}_2, \cdots \tilde{x}_n)$
for $k = 1 \ldots N$

* for each i we simulate;

$$\Delta x_i^{(k)} = \Delta_i \cdot \tan(\pi(\xi - 0.5)) \quad \Big\| \quad c_i^{(k)} = \tan(\pi - (\xi - 0.5))$$

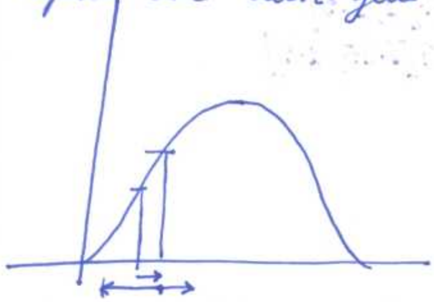
* compute $k_i = \max |\Delta x_i^{(k)}|$. || * compute $k = \max |c_i^{(k)}|$

* compute $\Delta x_i^{(k)} = \Delta_i \cdot \frac{c_i^{(k)}}{k}$

* $y^{(k)} = f(\tilde{x}_1 - \Delta x_1^{(k)}, \ldots \tilde{x}_n - \Delta x_n^{(k)})$

* $\Delta y^{(k)} = (\tilde{y} - y^{(k)}) \cdot k$.

Repeat the last few steps from approx. algo.



$$\frac{1}{\alpha}\left(f(x + \alpha \Delta x) - f(x)\right).$$

```
lower = 0.0;
upper = ...;
while (upper - lower > 0.01) {
    midpoint = (lower + upper)/2;
    f = F(midpoint);
    if (f < n/2.0) lower = midpoint;
    else {upper = mid;}
}
```

```
F()
{ sum = 0.0;
  for (k=1; k<1; k++)
  sum += 1.0((1 + Deltay[k]*Deltay[k])/
             (midpoint*midpoint));
}.
```

eg1:

$$\frac{1}{1+\left(\frac{\Delta y}{\Delta}\right)^2} = 2.$$

$$\frac{1}{1+\frac{\Delta y^2}{\Delta^2}} = 2 \Rightarrow 1+\frac{\Delta y^2}{\Delta^2} = 2 \Rightarrow \Delta y^2 = \Delta^2.$$

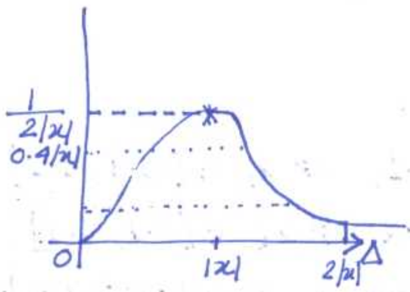
$$\therefore \Delta = |\Delta y|.$$

eg2:

$$\frac{1}{\pi\Delta}\cdot\frac{1}{1+\left(\frac{x}{\Delta}\right)^2} = \underset{max}{\rho(x)}.$$

$$\Delta = |x|$$

$$\frac{1}{2|x|}.$$



Richard Tapia.
Rice University.
3 pm: BUSN 319
Friday Oct 24th.

Inverse Problem as ;
Newton's Method.

HW:

$$\Delta = \sum_{i=1}^{N} \left| f(\tilde{x}_1, \dots \tilde{x}_{i-1}, \tilde{x}_i + \Delta_i, \dots \tilde{x}_n) - \tilde{y} \right|.$$

eg: $f(x_1, x_2) = x_1 - x_2$. $\qquad \tilde{y} = 0$.

$\tilde{x}_1 = 1.0$, $\tilde{x}_2 = 1.0$, $\Delta_1 = 0.1$, $\Delta_2 = 0.1$.

$x_1 \in [0.9, 1.1]$, $x_2 \in [0.9, 1.1]$.

$y = [-0.2, 0.2]$.

$$|f(\tilde{x}_1+\Delta, \tilde{x}_2) - 0.0| + |f(\tilde{x}_1, \tilde{x}_2+\Delta) - 0.0| = \Delta.$$

Random Error:

- Data Fusion, $\tilde{x}_1 \cdots \tilde{x}_n \rightarrow \tilde{x} = \dfrac{\sigma_1^{-2} x_1 + \sigma_2^{-2} x_2 + \cdots}{\sigma_1^{-2} + \sigma_2^{-2} + \cdots}$

- Data Processing, several algorithms.

Interval Error: (only know upper bound, we don't know meas. error).

- Data Fusion.

∇ Data processing.

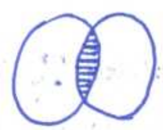
$\tilde{x}_1 = 1.0, \ \Delta_1 = 0.1, \Rightarrow x_1 \in [0.9, 1.1]$

$\tilde{x}_2 = 0.9, \ \Delta_2 = 0.2, \Rightarrow x_2 \in [0.7, 1.1]$

$\tilde{x}_3 = 1.02, \ \Delta_3 = 0.005. \Rightarrow \cancel{x_3 \in [0.95, 1.08]} \ x_3 \in [1.02, 1.012]$

$\therefore x \in [1.02, 1.1]$ $\{$Intersection$\}$.

$\underbrace{1.02}_{\text{lower bound of max.}}$, $\underbrace{1.1}_{\text{upper bound of min.}}$



$[\tilde{x} - \Delta, \tilde{x} + \Delta]$

$\tilde{x} = \dfrac{\underline{x} + \overline{x}}{2}$ midpoint, $\Delta = \dfrac{\overline{x} - \underline{x}}{2}$ [half width]

$\therefore \tilde{x} = 1.06$, $\Delta = 0.04$.

Widths of $\tilde{x}$ are smaller/equal to individual widths.

Trust:

$\tilde{x}_1 = 1.0, \ \Delta = 0.1. \Rightarrow x \in [0.9, 1.1]$. We are not 100% sure of measurements.

$\tilde{x}_2 = 1.1, \ \Delta = 0.2. \Rightarrow x \in [0.7, 1.1]$

This is extremely important in the interval case:

$\tilde{x}_3 = 0.8, \ \Delta_3 = 0.05.$

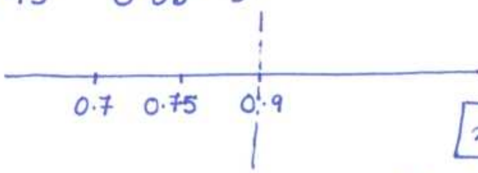
Outlier:

We conclude $\underbrace{x \in [0.75, 0.85]}_{\text{empty intersection} \Rightarrow \emptyset}$.

So one of these values is wrong.

$$\left.\begin{array}{lll} 1) & 0.9 & 1.1 \\ 2) & 0.7 & 1.1 \\ 3) & 0.75 & 0.85 \end{array}\right\}$$

Case: Two measurements are correct out of 3

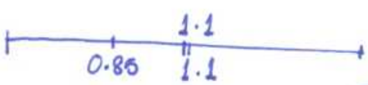


[If all 3 were correct]

$\boxed{x \geqslant 0.75}$

$$\left.\begin{array}{l} \left.\begin{array}{l}\to 0.9 \text{ is correct lower bound}\end{array}\right\} \Rightarrow x \geqslant 0.9 \\ \left.\begin{array}{l}\to 0.9 \text{ is a fluke not a correct lower bound}\end{array}\right\} \Rightarrow x \geqslant 0.75. \end{array}\right\} \Rightarrow x \geqslant 0.75.$$

So if certain no of values are incorrect, we sort all the values & eliminate the upper values.



$$\left.\begin{array}{l} \to x \leqslant 0.85 \\ \text{or} \\ \to x \leqslant 1.1 \end{array}\right\} \Rightarrow x \leqslant 1.1.$$

$\therefore x \in [0.75, 1.1]$.

This is called <u>fusing of interval data</u>:

Algorithm:

Lower Bound: $\tilde{x}_i \quad \Delta_i \qquad (p)$ — probability of ~~trust~~ correctness (trust / reliability).

$1 \leqslant i \leqslant n$.

$x_1 = \tilde{x}_1 - \Delta_1, \; \ldots \ldots \; x_n = \tilde{x}_n - \Delta_n$.

Sort them in increasing order:

$\underline{x}_{(1)} \leqslant \underline{x}_{(2)} \leqslant \underline{x}_{(3)} \cdots \underline{x}_{(n-1)} \leqslant \underline{x}_{(n)}$

we take;

$\underline{x}(n.p) \qquad n.(1-p)$

Eg:

$x_1 = 0.9, \; x_2 = 0.7, \; x_3 = 0.75.$

$x_{(1)} = 0.7 \leqslant x_{(2)} = 0.75 \leqslant x_{(3)} = 0.9.$

$(1) = 2 \quad (2) = 3 \quad (3) = 4.$

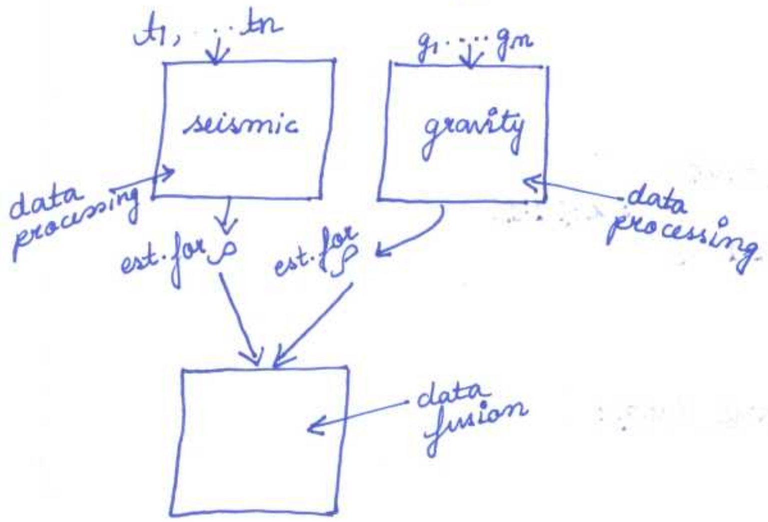
Upper Bound:

$\tilde{x}_i = \tilde{x}_i + \Delta_i$

$\overline{x}_{(1)} \leqslant \overline{x}_{(2)} \leqslant \cdots \leqslant \overline{x}_{(n)}$

$\overline{x}(n.(1-p))$.

Relation to Cyber Infrastructure:



Trust:

∀ data fusion.

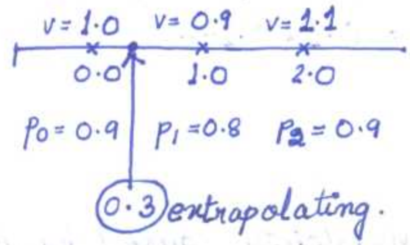
data processing. $\overset{*}{0.0}$ $*\,0.1$ $\overset{*}{0.2}$

Toy examples.

motivations: meas   Simplest ways: K-NN.

1-NN

eg:

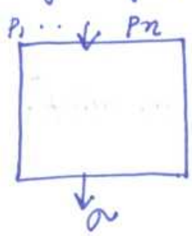


Everything is perfect; ⇒ 1.0 (predicted value), prob = 0.9.

if first is wrong & 2nd is correct ⇒ 0.9, prob = $(1-0.9) * (0.8) = 0.08$

if " " " " wrong ⇒ 1.1, " = 0.018.

Everything is wrong ⇒ ?, " = 0.002 → nothing.



Trust → uncertainity.

Measured Value:

is not exactly correct: ① (uncertainty) → Actual value is slightly diff.
② (trust) → Actual value is very diff.

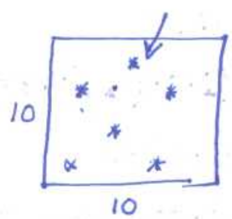
Standard way to describe spread: standard deviation

$$E[x] = p_1 x_1 + p_2 x_2 + p_3 x_3$$
$$= 0.9 \times 1.0 + 0.08 \times 0.9 + 0.02 \times 1.1$$
$$= 0.9 + 0.072 + 0.022 = 0.994.$$

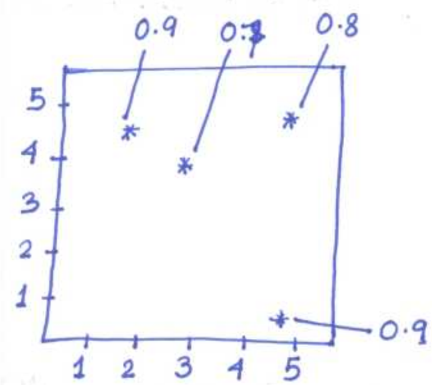
$$V[x] = p_1 \cdot (x_1 - E)^2 + p_2 (x_2 - E)^2 + p_3 (x_3 - E)^2$$
$$= (0.9) \times (0.006)^2 + (0.08) \times (0.094)^2 + (0.02) \times (0.106)^2 \ldots$$
$$= 0 \times 0.9 + 0.08 \cdot 0.1^2 + 0.02 \times (0.1)^2 = 0.1 \times 0.1^2$$
$$\therefore \sigma = \sqrt{V} = 0.03.$$

H/W:

1) program (due Nov 4). 2-NN and estimate uncertainty caused by trust.

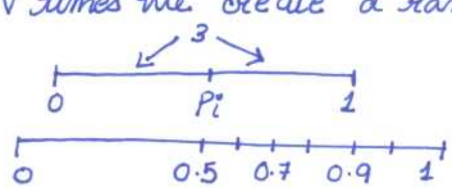


know: x_i y_i u_i p_i
want: extrapolate values on a grid 10×10.
first: extrapolate values.
then: find accuracy by Monte Carlo Simulations.



Algorithm:

N times we create a "random map".



0.7 0.8 0.65 0.7 ...

$$\sqrt{\frac{1}{n} \sum (x_i - a)^2}$$

N times.

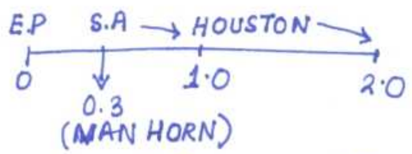
$\{$ – create a random map by picking ith point w/ prob p_i;
– based on this random map, we do extrapolation 2-NN

for each point on the grid, you get N interpolated values.

$$V^{(1)} \quad V^{(2)} \ \ldots \ V^{(n)}$$

$$mean = \frac{V^{(1)} + V^{(2)} + \cdots V^{(n)}}{n}.$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{k=1}^{N} \left(V^{(k)} - mean\right)}.$$



[In 1 dimension].

$x_1 = 0.0 \quad V_1 = 1.0 \quad p_1 = 0.9$

$x_2 = 1.0 \quad V_2 = 0.9 \quad p_2 = 0.8$

$x_3 = 2.0 \quad V_3 = 1.1 \quad p_3 = 0.9$

N = 100 times → 90 times El Paso, 8 times SA, 2 times Houston.

$$\frac{\overbrace{1.0 + \cdots 1.0}^{90} + \overbrace{0.9 + \cdots 0.9}^{8} + \cancel{0}\cdot \overbrace{1.1 + 1.1 + \cdots 1.1}}{100}$$

Same eg: 2NN

(1,2) $0.9 \times 0.82 = 0.72$ | $\frac{1.0 + 0.9}{2} = 0.95$

(1,3) $0.9 \times 0.2 \times 0.9 = 0.81 \times 0.2 = 0.162$ | $\frac{1.0 + 1.1}{2} = 1.05.$

(2,3) $0.1 \times 0.8 \times 0.9 = 0.072$ | $\frac{0.9 + 1.1}{2} = 1.0$

0.954

prob that only 1 point remains ← $\boxed{0.046}$

How do we keep w/. prob p?

- create $\xi \in$ Unif $[0, 1]$
- if $\xi \leq p$ we keep
  else we don't keep.

```
for (i=0; i<n; i++)
  { if (random() < p[i])
        keep[i] = true;
    else
        keep[i] = false;
  }
```