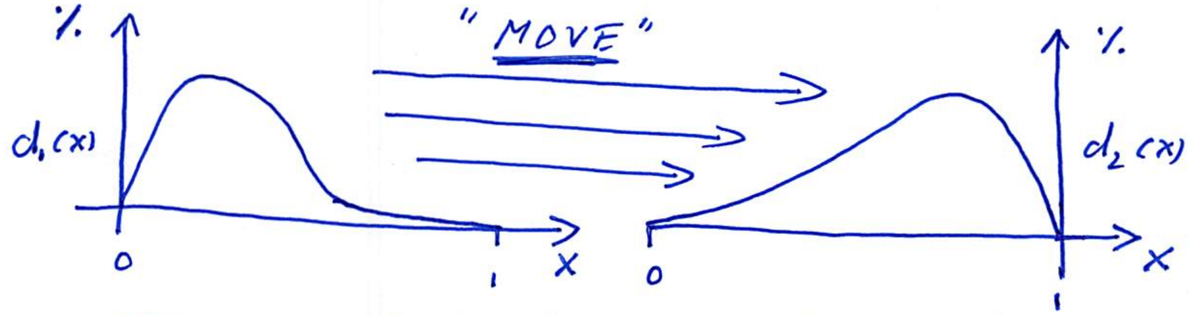
Bernd Hamann
Stratovan

3/7/2016

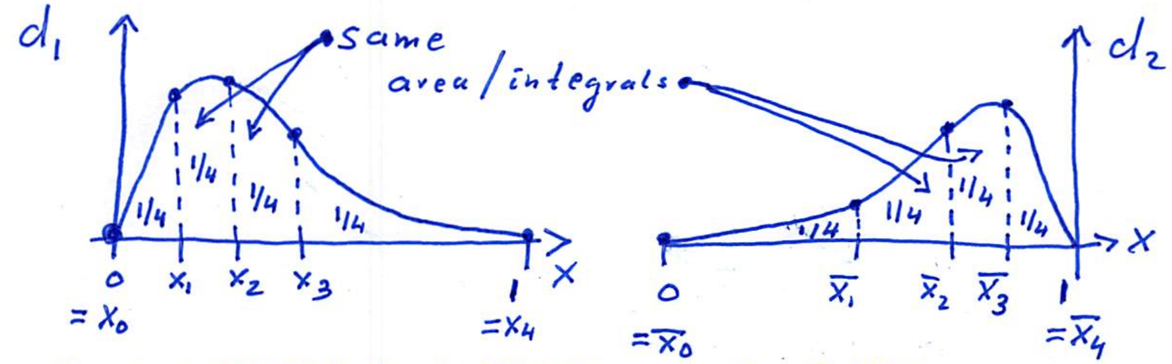
A Measure for Defining the Distance between 2 Histograms / Distributions – A CUBIC SPLINE-based Approach

- IDEA: "Moving a Distribution to another Distribution"



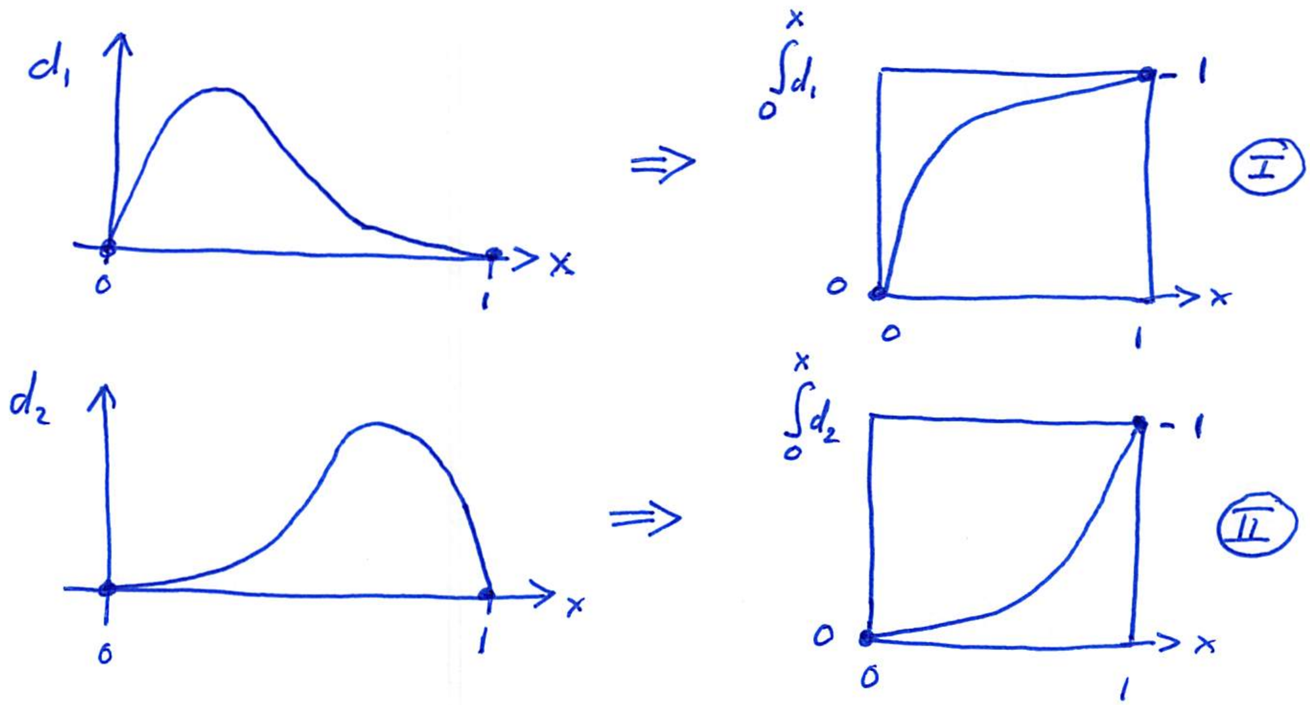
- The 2 probability density functions / distributions are both normalized, i.e., $\int_0^1 d_1(x)\,dx = \int_0^1 d_2(x) = \underline{\underline{1}}$.

- Determine x-values x_i and $\overline{x}_i$ such that the total area under the curves of d_1 and d_2 are uniformly distributed over intervals $[x_i, x_{i+1}]$ and $[\overline{x}_i, \overline{x}_{i+1}]$, respectively:

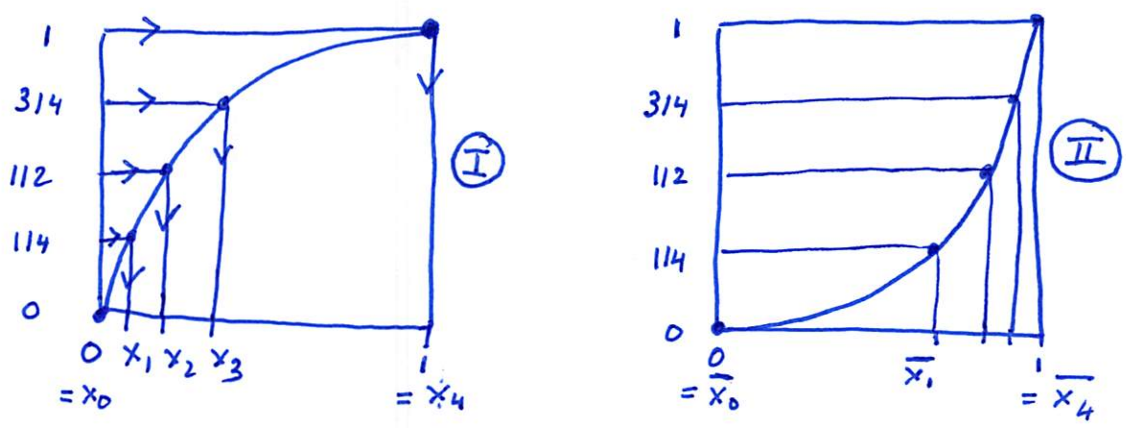


! $\Rightarrow$ ESTABLISH INTEGRAL-BASED CORRESPONDENCE BETWEEN d_1 and d_2: $x_i \longleftrightarrow \overline{x}_i$, $i = 0 \dots k$. **!**

- In other words, use <u>accumulated integral values</u> to determine the "<u>split points</u>" x_i and $\overline{x}_i$ for d_1 and d_2:



- <u>Split point generation</u>:



$\Rightarrow$ 2 split point sets defining correspondence: $\{x_i\}_{i=0}^{K}$, $\{\overline{x}_i\}_{i=0}^{K}$

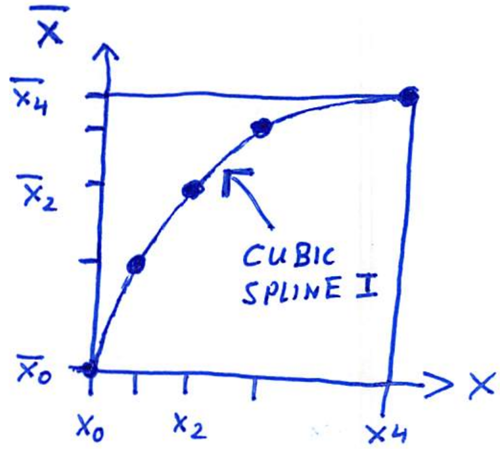
"<u>x_i corresponds to $\overline{x}_i$</u>"

Bernd Hamann
Stratovan

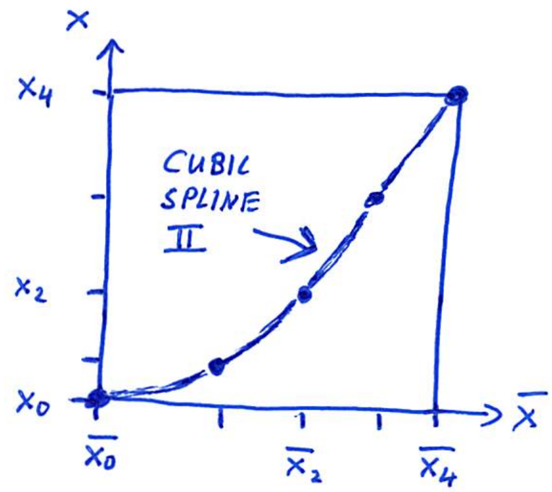
3/7/2016

- **Next**: Use the **NATURAL CUBIC SPLINE** (energy-minimizing!)

to map x_i to $\overline{x}_i$ and $\overline{x}_i$ to x_i, $i = 0 \ldots K$:



$x \mapsto \overline{x}$



$\overline{x} \mapsto x$

$\Rightarrow$ 2 splines result: $s_1 = \overline{x}(x)$,
("natural end conditions") $s_2 = x(\overline{x})$

$\Rightarrow$ Can compute **ENERGY** for these splines:

$$E_1 = \int_0^1 (s_1'')^2 \, dx \quad ,$$

$$E_2 = \int_0^1 (s_2'')^2 \, d\overline{x} \quad .$$

!! $\Rightarrow$ **DEFINE**: **DISTANCE** (d_1, d_2)

$$= \frac{1}{2}(E_1 + E_2) \quad .$$

(Advantage over TPS-based method:
- No need for corresponding "landmarks"
- Simple, efficient, robust method)

~ BH