Sequential Detection
Overview & Open Problems

George V. Moustakides,
University of Patras, GREECE

Outline

- Sequential hypothesis testing
  - SPRT
  - Application from databases
  - Open problems
- Sequential detection of changes
  - The Shiryaev test
  - The Shiryaev-Roberts test
  - The CUSUM test
  - Open problems
  - Decentralized detection (sensor networks)

Sequential Hypothesis testing

Conventional binary hypothesis testing:

Fixed sample size observation vector $X=[x_1,\ldots,x_K]$

X satisfies the following two hypotheses:

$$\mathsf{H}_0 : \; X \sim f_0(X)$$
$$\mathsf{H}_1 : \; X \sim f_1(X)$$

Given the data vector X, decide between the two hypotheses.

Decision rule: $\mathsf{D}(X) \in \{0,1\}$

Bayesian formulation

$$\min_{\mathsf{D}}\{\mathsf{P}(\mathsf{H}_0)\mathsf{P}(\mathsf{D}=1|\mathsf{H}_0)+\mathsf{P}(\mathsf{H}_1)\mathsf{P}(\mathsf{D}=0|\mathsf{H}_1)\}$$

Neyman-Pearson formulation

$$\min_{\mathsf{D}}\mathsf{P}(\mathsf{D}=0|\mathsf{H}_1); \quad \text{subject}: \mathsf{P}(\mathsf{D}=1|\mathsf{H}_0)\leq\alpha$$

Likelihood ratio test:

$$\frac{f_1(X)}{f_0(X)} \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtreqless}} \lambda$$

For i.i.d.:

$$u_K=\sum_{n=1}^{K}\log\left(\frac{f_1(x_n)}{f_0(x_n)}\right) \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtreqless}} \lambda'$$

Sequential binary hypothesis testing

Observations $x_1,\ldots,x_t,\ldots$ become available **sequentially**

$$\mathsf{H}_0: \; x_1,\ldots,x_t,\ldots \sim f_0(x_1,\ldots,x_t,\ldots)$$

$$\mathsf{H}_1: \; x_1,\ldots,x_t,\ldots \sim f_1(x_1,\ldots,x_t,\ldots)$$

Time	Observations	Decision
1	x_1	$\mathsf{D}(x_1)$
2	x_1, x_2	$\mathsf{D}(x_1, x_2)$
$\vdots$	$\vdots$	$\vdots$
t	$x_1, \ldots, x_t$	$\mathsf{D}(x_1, \ldots, x_t)$
$\vdots$	$\vdots$	$\vdots$

We apply a **two-rule** procedure

1st rule: at each time instant t, evaluates whether the observed data can lead to a reliable decision

Time	Observations
1	x_1
2	x_1, x_2
$\vdots$	$\vdots$
T	$x_1, \ldots, x_T$

STOP getting more data.

T is a **stopping rule** Random!

$$\mathrm{D}(x_1, \ldots, x_T) \in \{0, 1\}$$

2nd rule: Familiar decision rule

Why Sequential ?

On average, we need significantly less samples to reach a decision than the fixed sample size test, for the same level of confidence (same error probabilities)

For the Gaussian case it is 4 - 5 times less samples.

SPRT (Wald 1945)

$$u_t = \log\left(\frac{f_1(x_1,\ldots,x_t)}{f_0(x_1,\ldots,x_t)}\right)$$

$$u_t = \sum_{n=1}^{t} \log\left(\frac{f_1(x_n)}{f_0(x_n)}\right) = u_{t-1} + \log\left(\frac{f_1(x_t)}{f_0(x_t)}\right)$$

Here there are **two** thresholds $A < 0 < B$

Stopping rule: $T = \inf\{t > 0 : u_t \notin (A, B)\}$

Decision rule:

$$\mathrm{D}(x_1,\ldots,x_T) = \begin{cases} 1 & \text{when } u_T \geq B \\ 0 & \text{when } u_T \leq A \end{cases}$$


H1
B
Infinite Horizon
u_t
T
0
t
A
H0

Amazing optimality property!!!

$$\mathsf{P}(\mathsf{D}=1|\mathsf{H}_0) \le \alpha \qquad \mathsf{P}(\mathsf{D}=0|\mathsf{H}_1) \le \beta$$

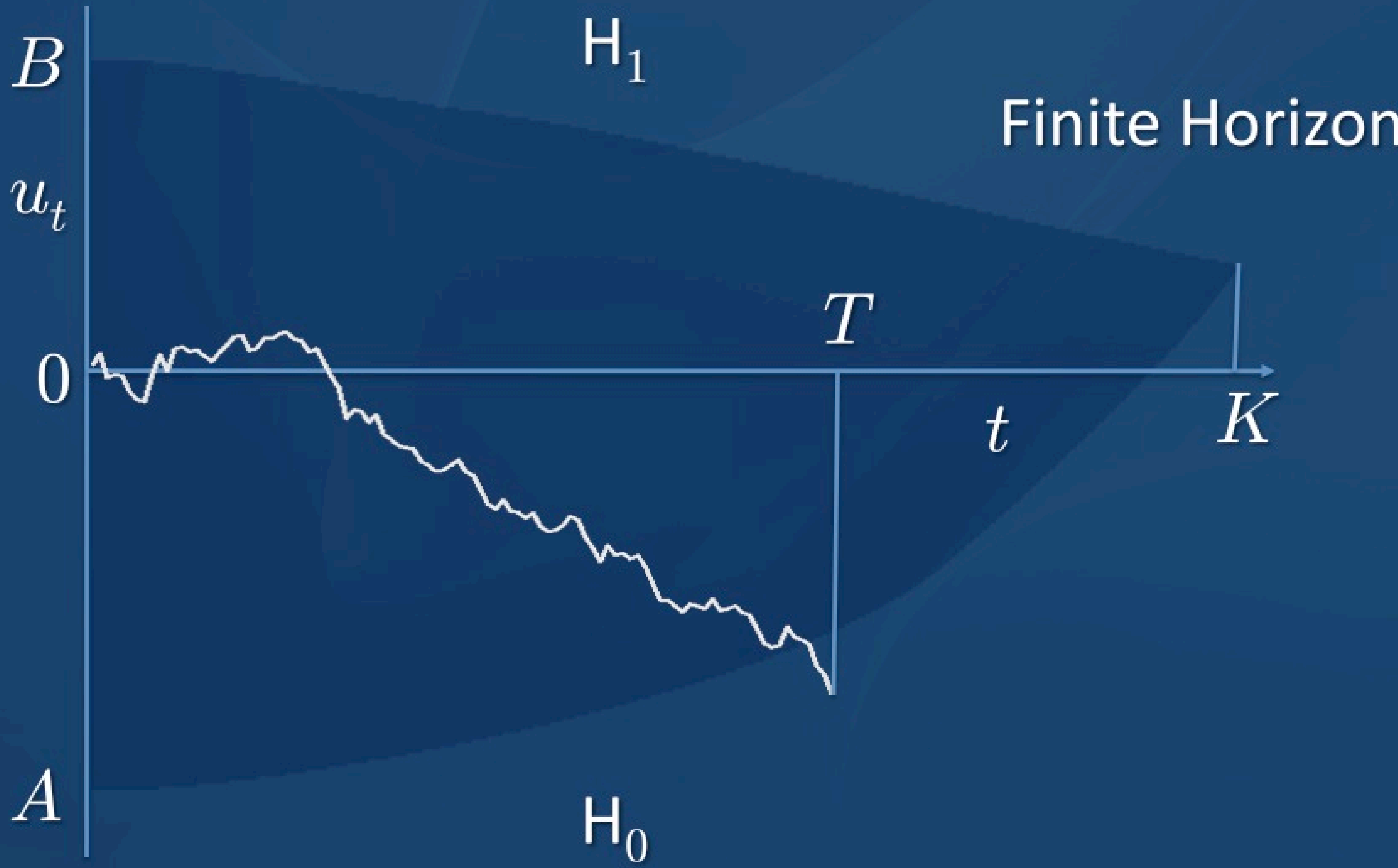
$$\min_{T,\mathsf{D}} \mathsf{E}[T|\mathsf{H}_0] \qquad \min_{T,\mathsf{D}} \mathsf{E}[T|\mathsf{H}_1]$$

SPRT solves both problems simultaneously

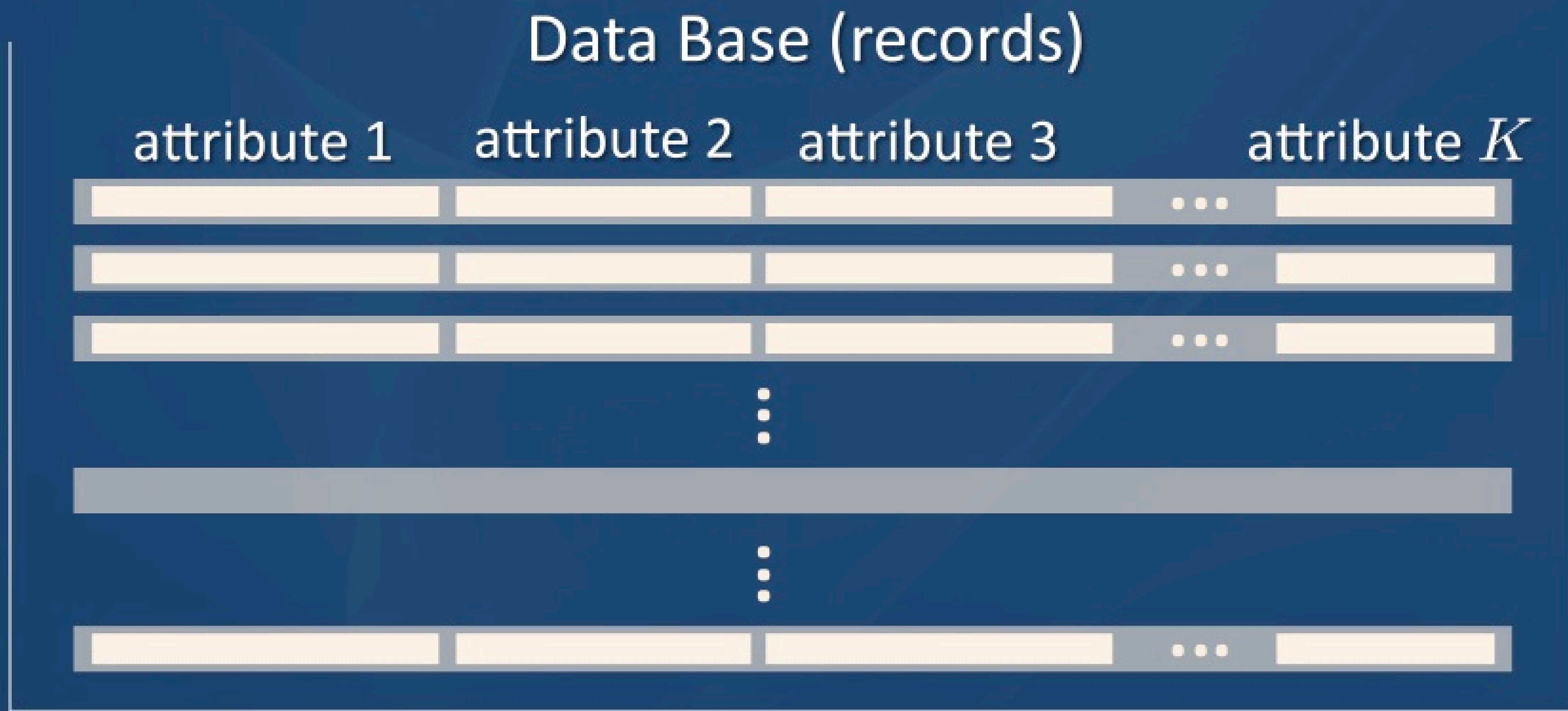
A, B need to be selected to satisfy the two error probability constraints with equality

- I.i.d. observations (1948, Wald-Wolfowitz)
- Brownian motion (1967, Shiryaev)
- Homogeneous Poisson (2000, Peskir-Shiryaev)

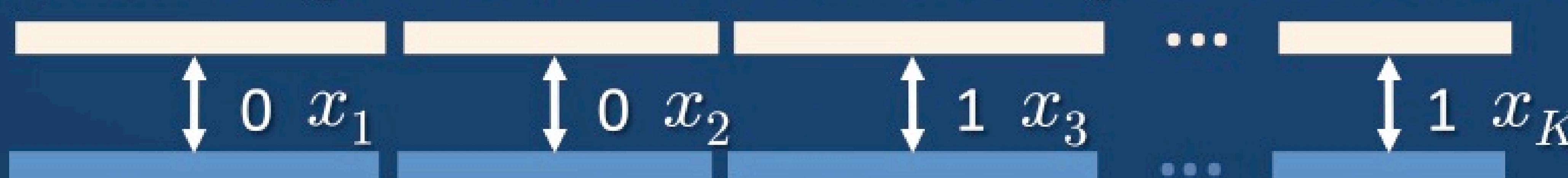
Open Problems: Dependency, Multiple Hypotheses

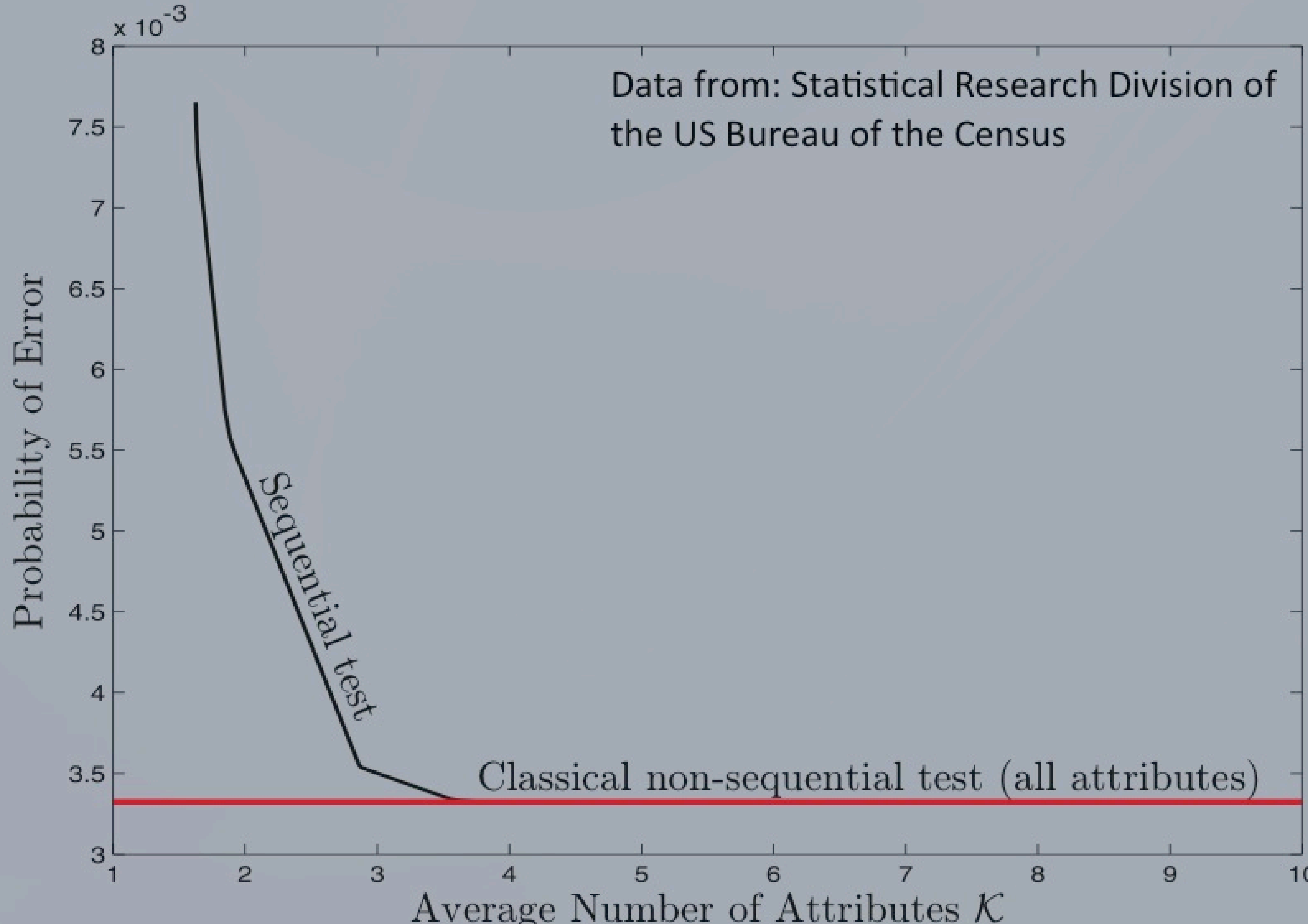

B
H_1
Finite Horizon
u_t
T
0
t
K
A
H_0

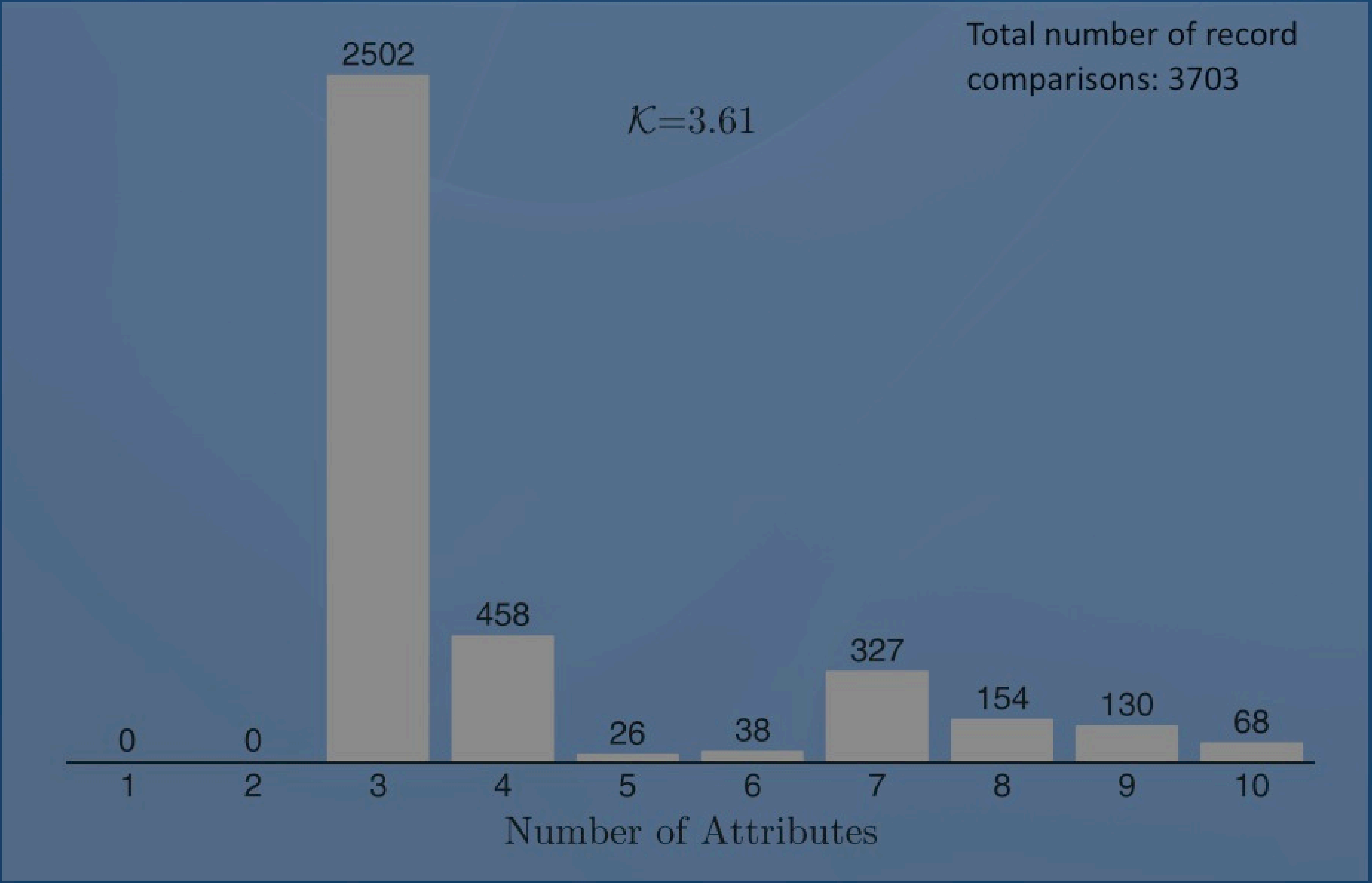
Record linkage



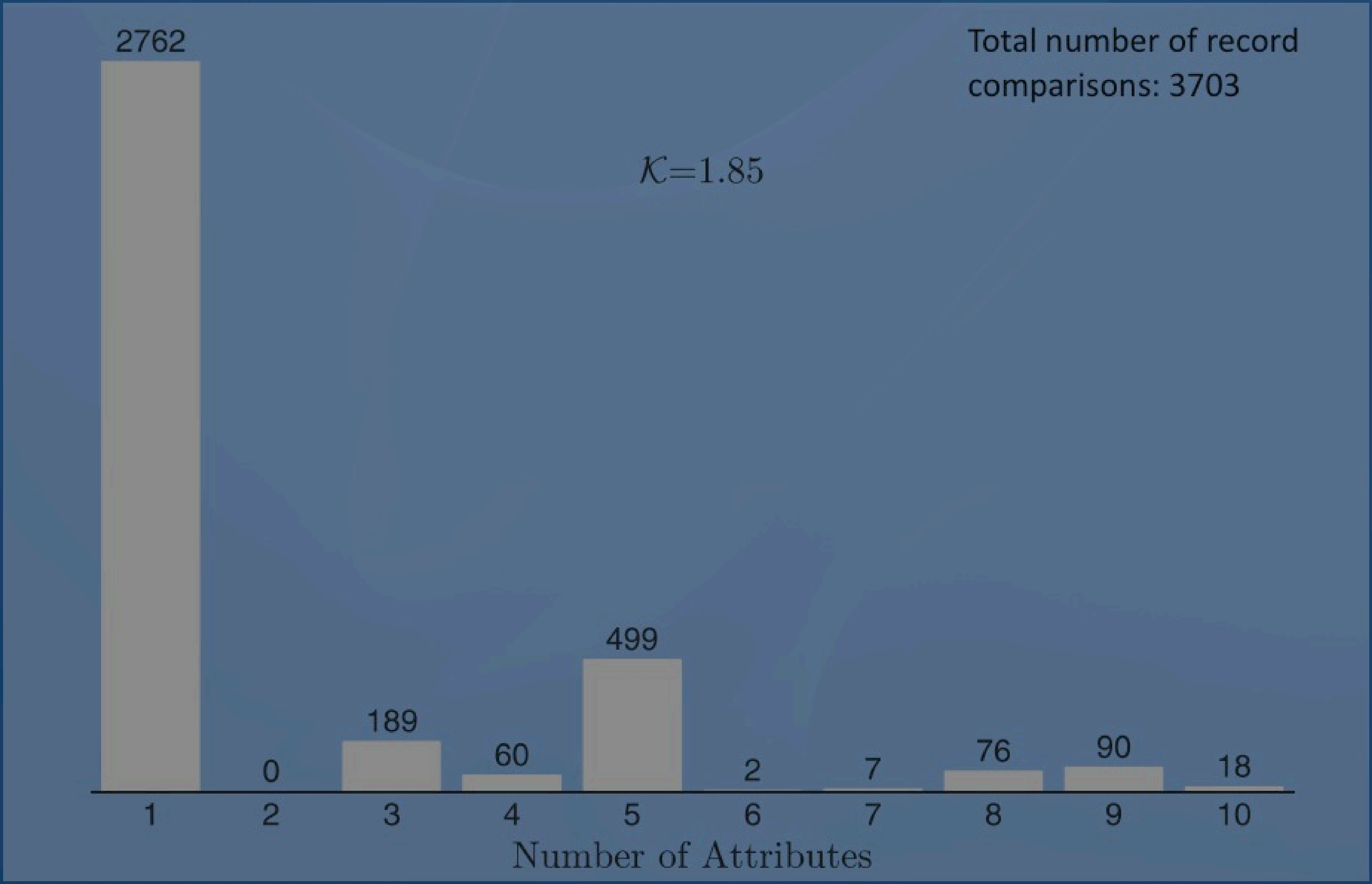
We assume known probabilities for x_i =0 or 1 under match (Hypothesis H_1) or nonmatch (Hypothesis H_0) for each attribute.

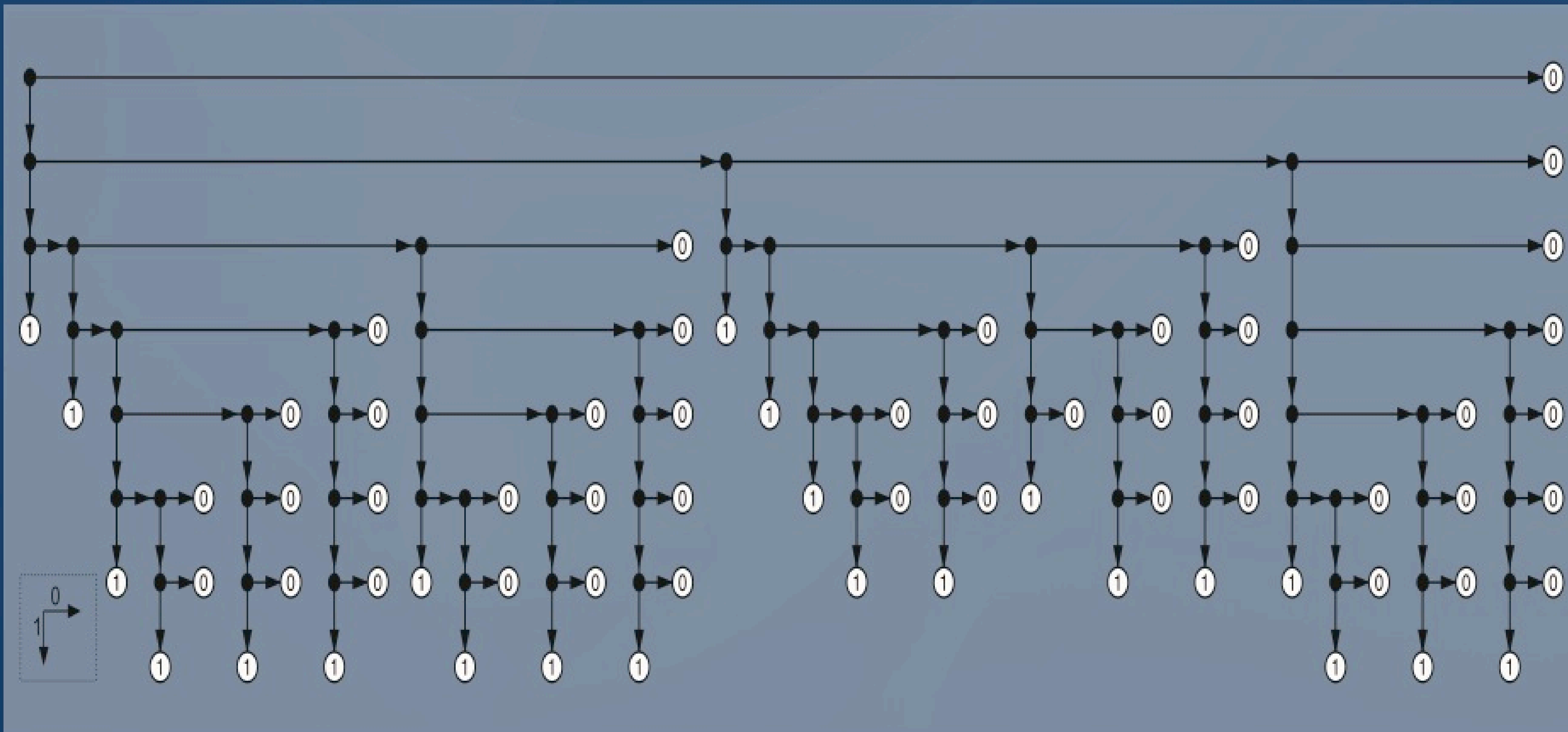



x 10^-3
8
7.5
7
6.5
6
5.5
5
4.5
4
3.5
3
1
2
3
4
5
6
7
8
9
10
Probability of Error
Average Number of Attributes $\mathcal{K}$
Data from: Statistical Research Division of the US Bureau of the Census
Sequential test
Classical non-sequential test (all attributes)



Total number of record comparisons: 3703
$\mathcal{K}$=3.61
2502
458
327
154
130
68
26
38
0
0
1
2
3
4
5
6
7
8
9
10
Number of Attributes

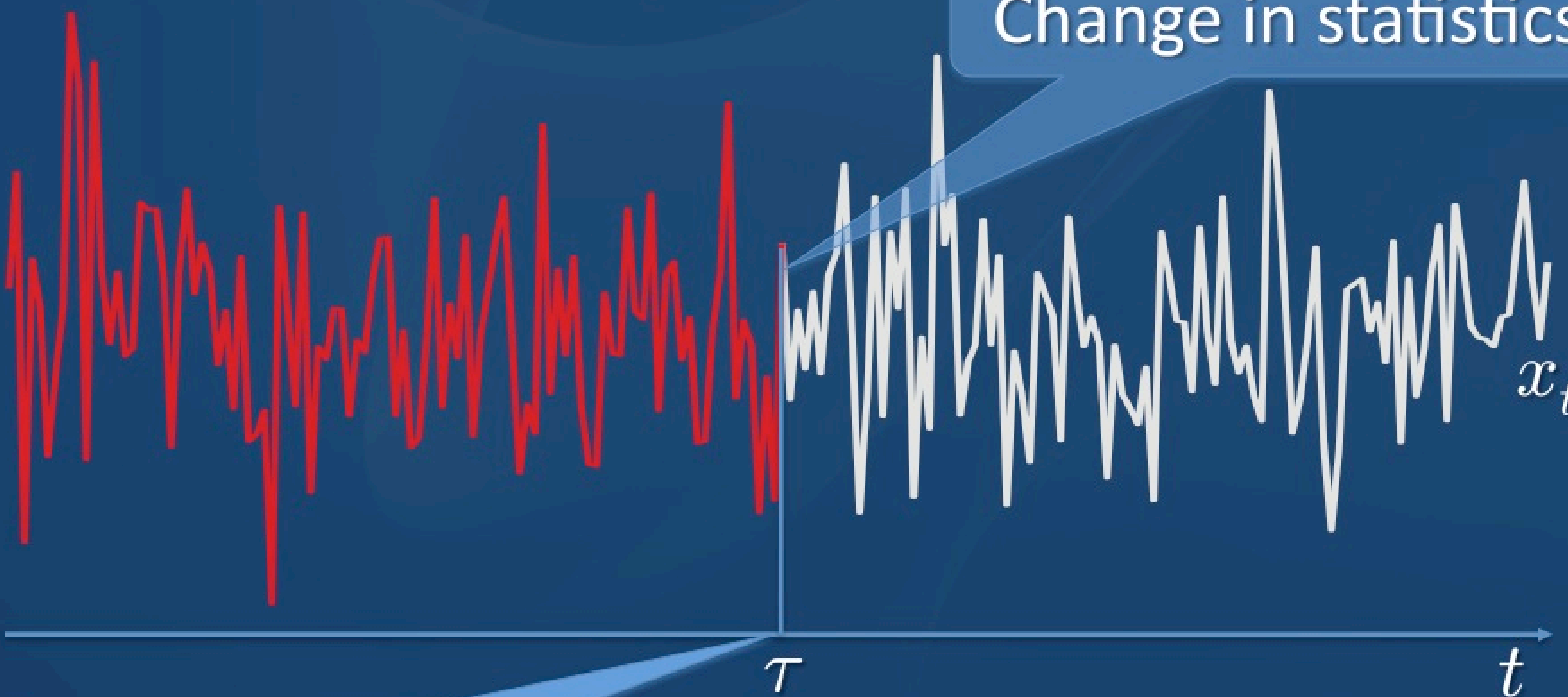

2762
0
189
60
499
2
7
76
90
18
1
2
3
4
5
6
7
8
9
10
Number of Attributes
$\mathcal{K}$=1.85
Total number of record comparisons: 3703



In collaboration with V. Verykios

Sequential change detection

Change in statistics



Detect occurrence
as soon as possible

Applications

Monitoring of quality of manufacturing process (1930's)

Biomedical Engineering

Electronic Communications

Econometrics

Seismology

Speech & Image Processing

Vibration monitoring

Security monitoring (fraud detection)

Spectrum monitoring

Scene monitoring

Network monitoring and diagnostics (router failures, intruder detection)

Databases .....

Mathematical setup

We observe **sequentially** a process $\{x_t\}$ that has the following statistical properties

$$x_t \sim \begin{cases} f_0 & \text{for } 0 \le t \le \tau \\ f_1 & \text{for } \tau < t. \end{cases}$$

- Changetime τ : either random with known prior or deterministic but unknown.
- Both pdfs f_0, f_1 are considered known.

Detect occurrence of τ as soon as possible

We are interested in **sequential** detection schemes.

Whatever sequential scheme one can think of, at every time instant t it will have to make one of the following two decisions:

- Either decide that a change didn't take place before t, therefore it needs to continue taking more data.
- Or that a change took place before t and therefore it should stop and issue an alarm.

Sequential Detector ←→ Stopping rule

Shiryaev test (Bayesian, Shiryaev 1963)

Changetime τ is random with Geometric prior.

If T is a stopping rule then we define the following cost function

$$J(T) = c\,\mathsf{E}[(T-\tau)^+] + \mathsf{P}(T \leq \tau)$$

Optimum T:

$$\min_T J(T)$$

Define the statistic: $\pi_t = \mathsf{P}(\tau \leq t | x_1, \ldots, x_t)$

There exists $\nu \in (0,1)$ such that the following rule is optimum.

$$T_S = \inf\{t > 0 : \pi_t \geq \nu\}$$

In discrete time when $\{x_t\}$ are i.i.d. before and after the change.

In continuous time when $\{x_t\}$ is a Brownian motion with constant drift before and after the change.

In continuous time when $\{x_t\}$ is Poisson with constant rate before and after the change.

Shiryaev-Roberts test (Minmax, Pollak 1985)

Changetime τ is deterministic and unknown.

For any stopping rule T define the following criterion:

$$J(T) = \sup_{\tau} \mathsf{E}_1[(T-\tau) | T > \tau]$$

Optimum T:

$$\min_{T} J(T)$$

subject to $\mathsf{E}_0[T] \geq \gamma$

In discrete time, when data are i.i.d. before and after the change with pdfs f_0, f_1.

Compute recursively the following statistic:

$$S_t = (1 + S_{t-1})\frac{f_1(x_t)}{f_0(x_t)};$$

Pollak (1985)

$$T_P = \inf\{t > 0 : S_t \geq \nu\}$$

$$[J(T_P) - \min_T J(T)] = o(1); \quad \text{as } \gamma \to \infty$$

~~Yakir (AoS 1997) provided a proof of strict optimality.~~
Mei (2006) showed that the proof was problematic.
Tartakovsky (2011) found a counterexample.

CUSUM test (minmax, Lorden 1971)

Changetime τ is deterministic and unknown.

For any stopping rule T define the following criterion:

$$J(T) = \sup_{\tau} \sup_{x_1,\ldots,x_\tau} \mathsf{E}_1[(T-\tau)^+|x_1,\ldots,x_\tau]$$

Optimum T:

$$\min_T J(T)$$

subject to $\mathsf{E}_0[T] \geq \gamma$

Compute the Cumulative Sum (CUSUM) statistic y_t as follows:

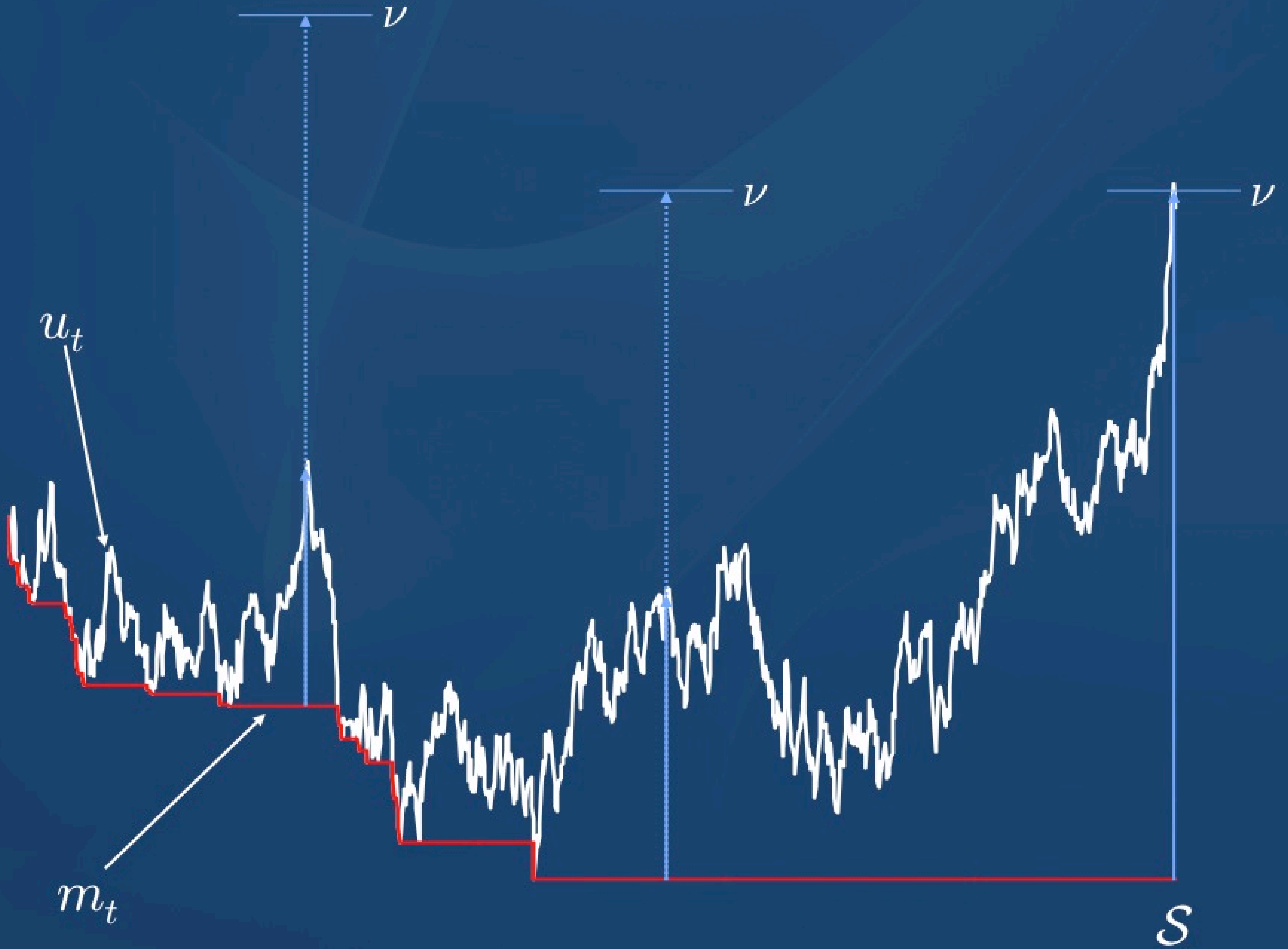
Running LLR $$u_t = \log\left(\frac{f_1(x_1,\ldots,x_t)}{f_0(x_1,\ldots,x_t)}\right)$$

Running minimum $$m_t = \inf_{0\leq s\leq t} u_s$$

CUSUM statistic $$y_t = u_t - m_{t-1}$$

CUSUM test $$\mathcal{S} = \inf\{t > 0 : y_t \geq \nu\}$$

For i.i.d. $$y_t = (y_{t-1})^+ + \log\left(\frac{f_1(x_t)}{f_0(x_t)}\right)$$



ν
ν
ν
u_t
m_t
$\mathcal{S}$

Discrete time: i.i.d. before and after the change

Lorden (1971) asymptotic optimality.

Moustakides (1986) strict optimality.

Continuous time

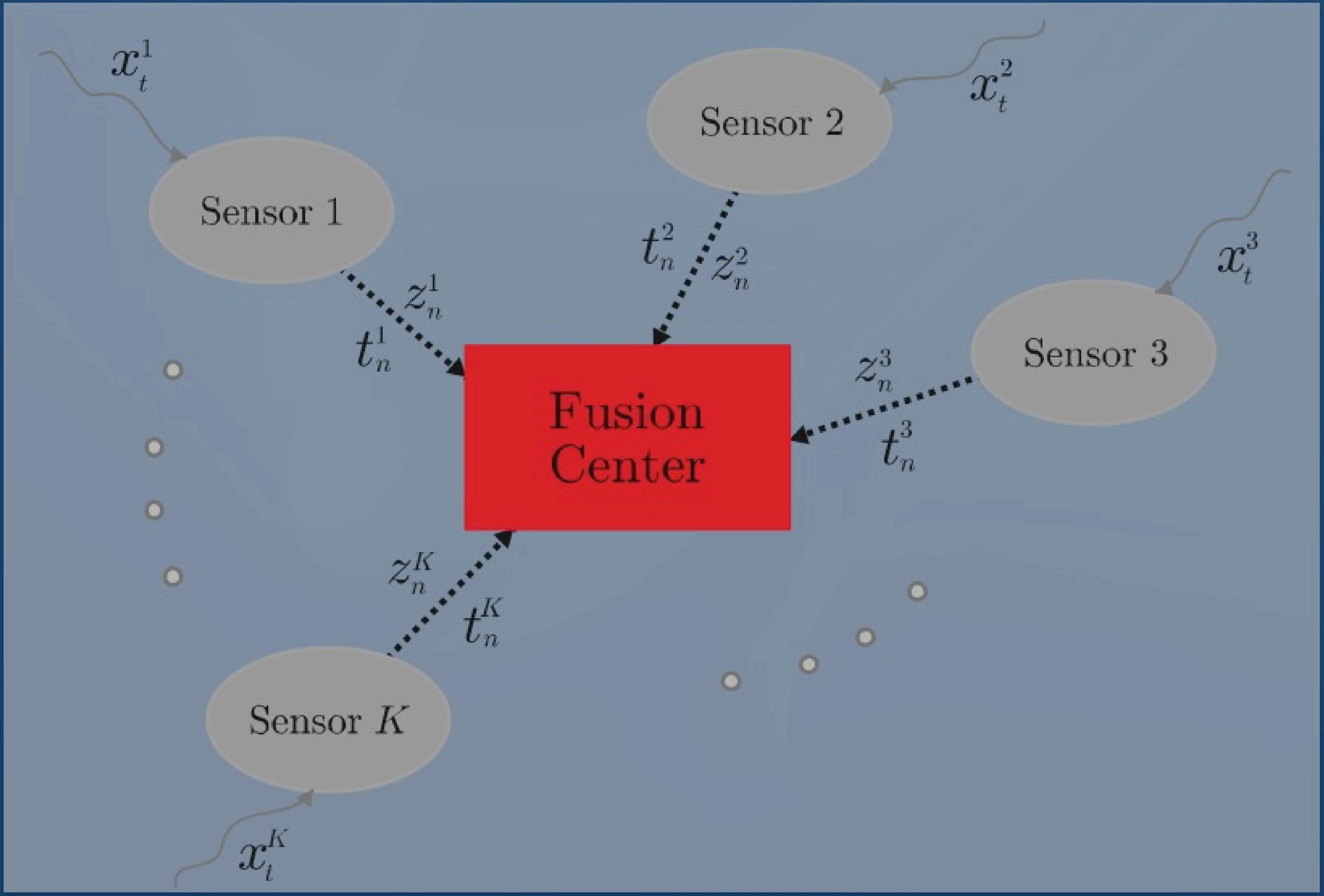
Shiryaev (1996), Beibel (1996) strict optimality for BM

Moustakides (2004) strict optimality for Ito processes

Moustakides (>2012) strict optimality for Poisson processes.

Open Problems: Dependent data, Non abrupt changes, Transient changes,…

Decentralized detection

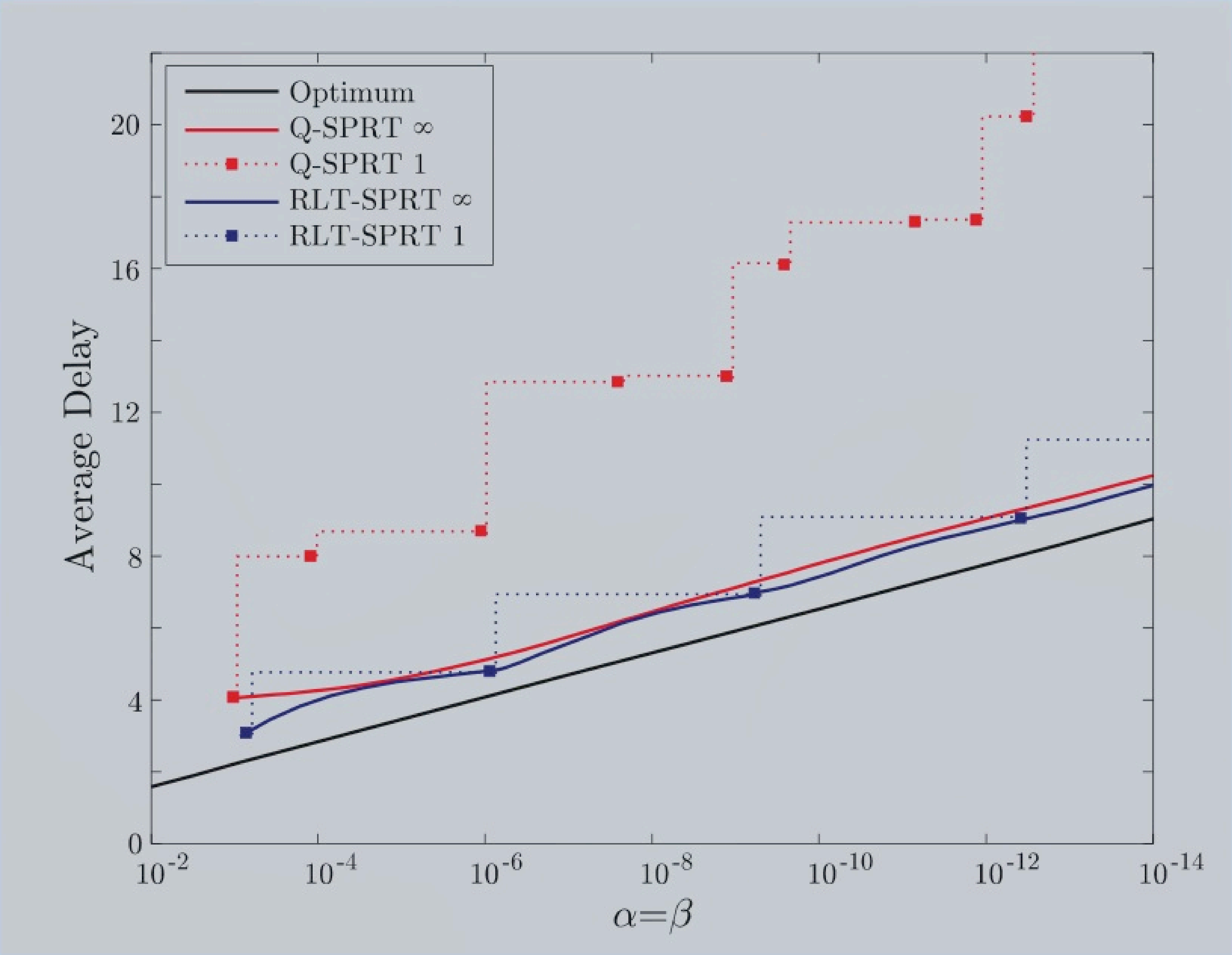


If more than 1 bits, **quantize overshoot!**

The Fusion center if at time t_n receives information from sensor i it updates an estimate of the global log-likelihood ratio:

$$\hat{u}_t = \begin{cases} \hat{u}_{t-} + B_i & \text{if bit is 1} \\ \hat{u}_{t-} + A_i & \text{if bit is 0} \end{cases}$$

and performs an SPRT (if hypothesis testing) or a CUSUM (if change detection) using the estimate of the global log-likelihood ratio.


Optimum
Q-SPRT ∞
Q-SPRT 1
RLT-SPRT ∞
RLT-SPRT 1
Average Delay
0
4
8
12
16
20
10-2
10-4
10-6
10-8
10-10
10-12
10-14
α=β