USC Viterbi School of Engineering

Group Testing for Efficient SARS-CoV-2 Assessment

Dhruva Kartik, Urbashi Mitra

Ashutosh Nayyar

Ming Hsieh Department of Electrical Engineering

University of Southern California

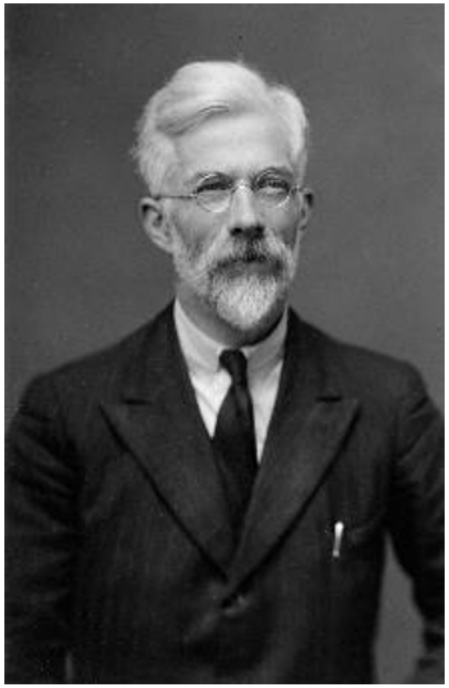
Peter Kuhn, Biological Sciences, Biomedical Engineering, Medicine

Neeraj Sood, Public Policy

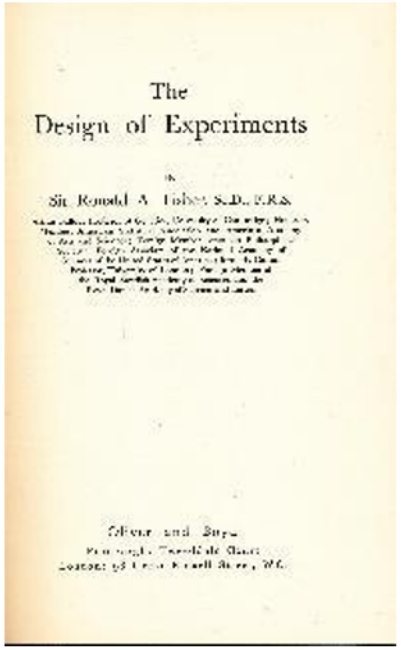
Thanks to: ONR N00014-15-1- 2550, NSF CNS-1213128, NSF CCF-1410009, ARO W911NF1910269, NSF CPS-1446901

Design of Experiments

USC Viterbi

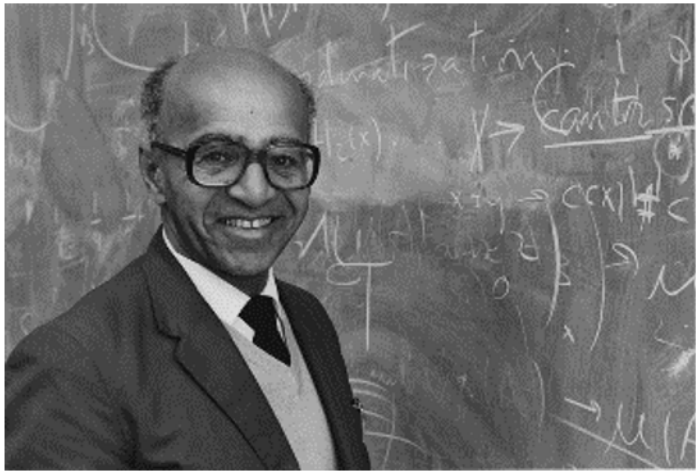


Sir Ronald Fisher
1890-1962

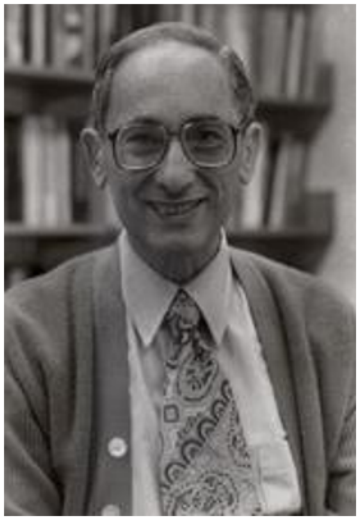

The
Design of Experiments
By
Sir Ronald A. Fisher Sc.D., F.R.S.

More Broadly

USC Viterbi



David Blackwell
1890-1962



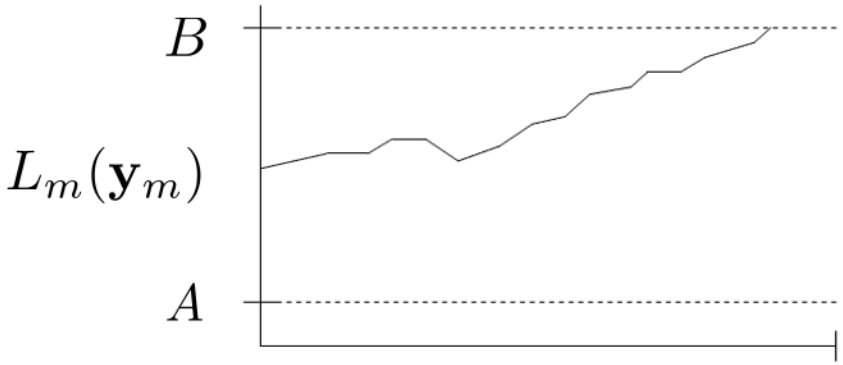
Herman Chernoff
1923-

Abraham Wald
1902-1950

Classical SPRT

USC Viterbi

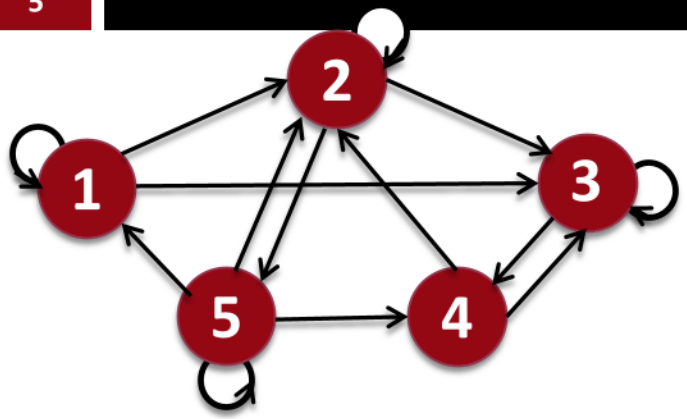
- Sequential probability ratio test
  - A Wald, *The Annals of Mathematical Statistics*, 1945
  - Samples: $\mathbf{y}_m = [y_1, y_2, \ldots, y_m]$
  - Likelihood ratio: $L_m(\mathbf{y}_m) = \dfrac{p_{\mathbf{y}_m|s_1}}{p_{\mathbf{y}_m|s_0}}$
  - Detection rule: $$\delta(L_m(\mathbf{y}_m)) = \begin{cases} s_0, & L_m(\mathbf{y}_m) \le A \\ s_1, & L_m(\mathbf{y}_m) \ge B \\ \text{sample}, & \text{else} \end{cases}$$

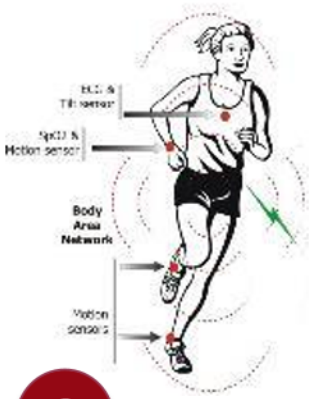


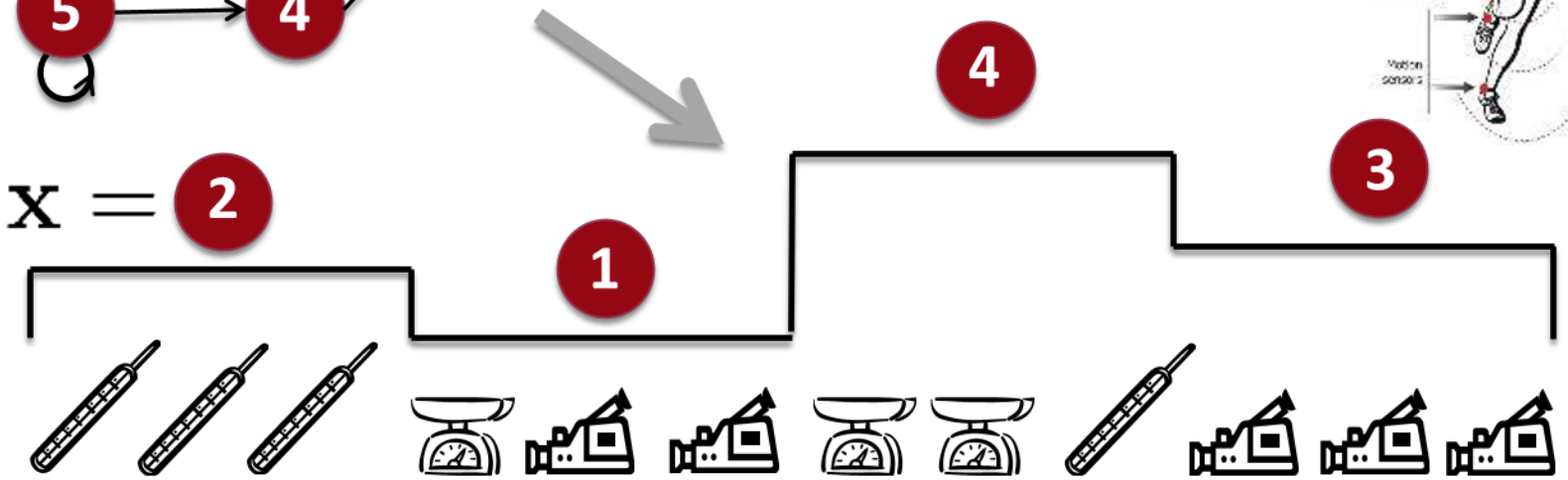
same experiment

Controlling Observations

USC Viterbi







$$\mathbf{y} = f(\mathbf{x}, \mathbf{u}) \rightarrow \hat{\mathbf{x}}$$

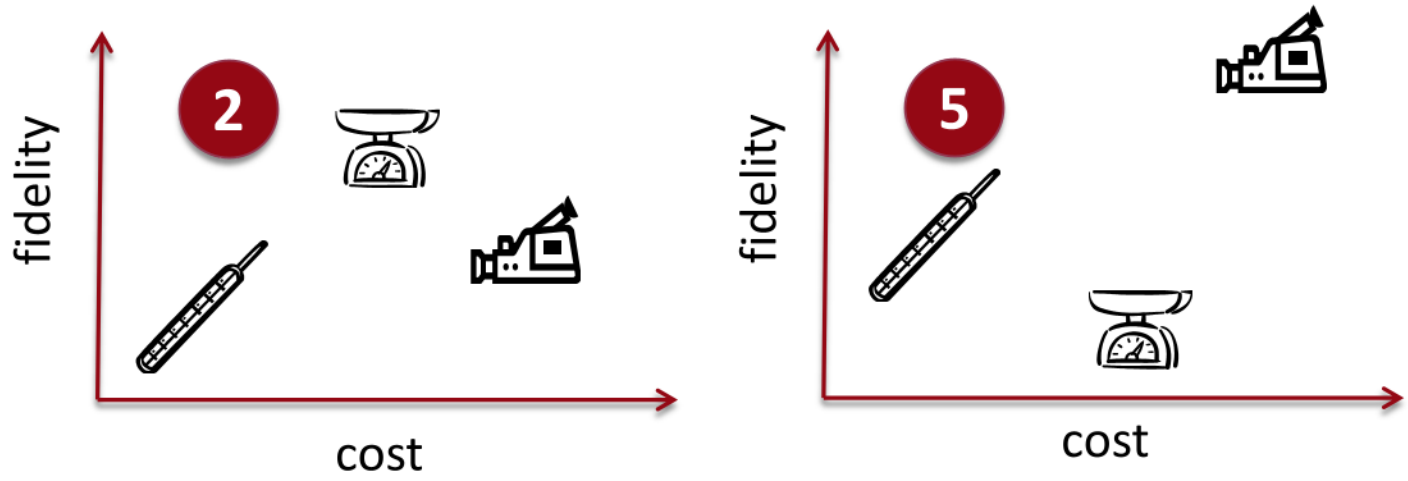
observation state, control

active hypothesis testing

Jovanov *et al. Journal of NeuroEngineering and Rehabilitation* 2005

Heterogeneity

USC Viterbi

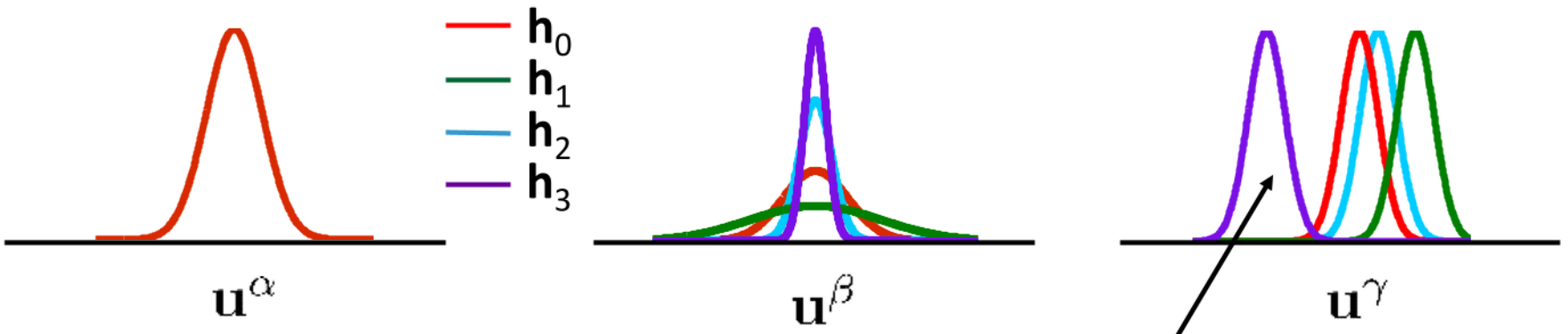


- Different sensors are good at discriminating different states
- True state influences best experiment/observations

The quality of observations

USC Viterbi

- How to quantify informativeness?



$\mathbf{u}^{\alpha}$ control = choice of experiment

likelihood of a particular hypothesis

- Choice of control makes hypotheses easier to distinguish

Metrics for distributions

USC Viterbi

- Relative entropy (Kullback-Leibler distance)

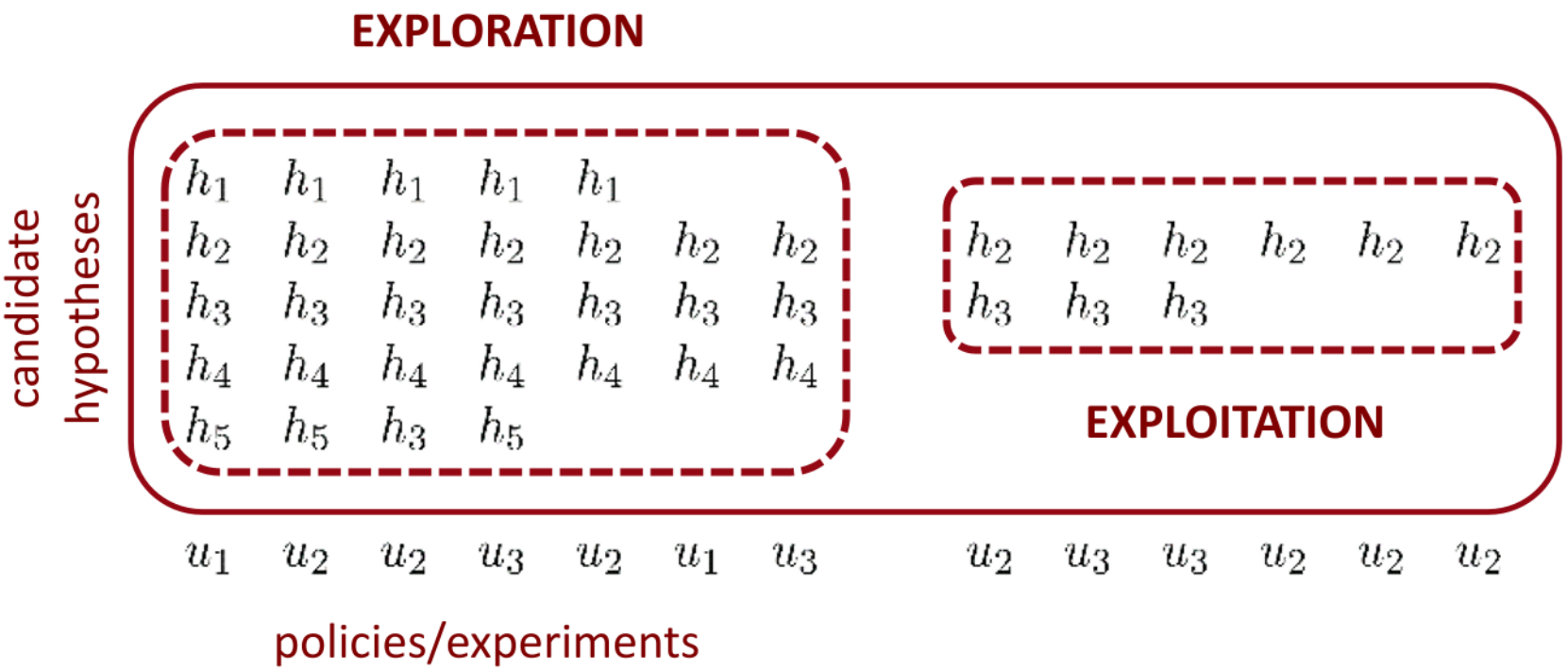
$$
\begin{aligned}
D(p||q) &= \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)} \\
&\geq 0 \\
D(p||q) &\neq D(q||p)
\end{aligned}
$$

- Not a true distance – does not satisfy triangle inequality, asymmetry…

S. Kullback & RA Leibler, "On information and sufficiency," *Annals of Mathematical Statistics,* 22(1): 79-86, 1951

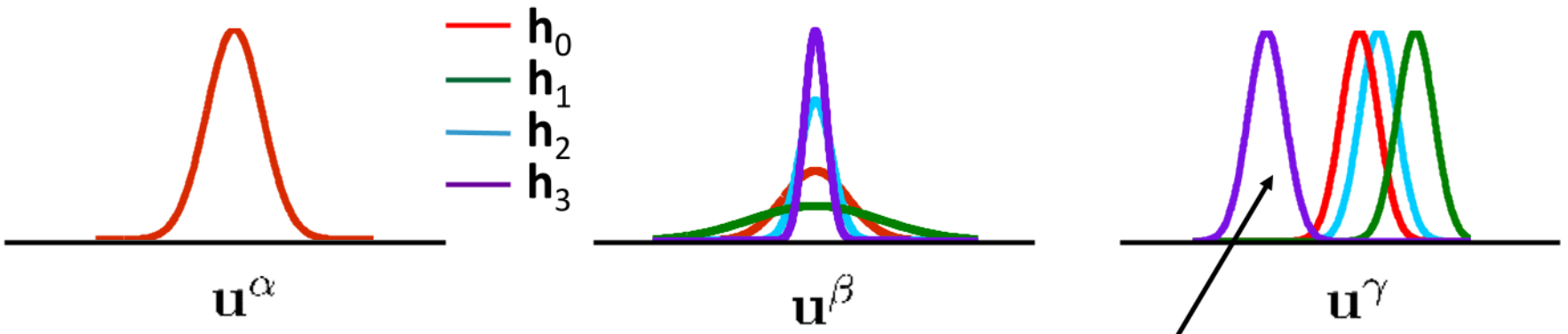
Active Hypothesis Testing

USC Viterbi



The quality of observations

USC Viterbi



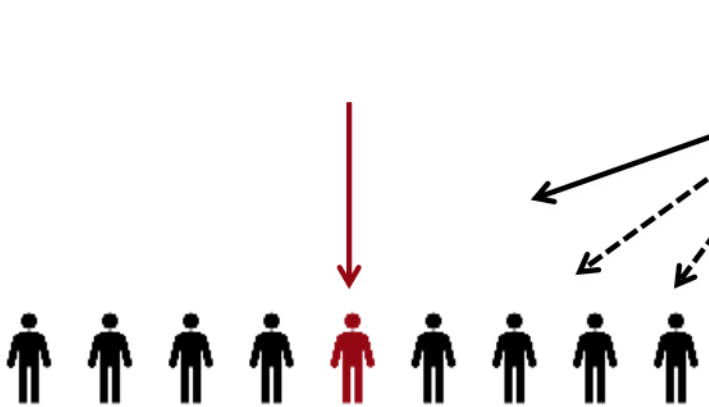
$\mathbf{u}^{\alpha}$ control = choice of experiment

likelihood of a particular hypothesis

- The most informative experiment depends on the true hypothesis

System Model

USC Viterbi



controller decides who to test at each time

one test for each person

M people

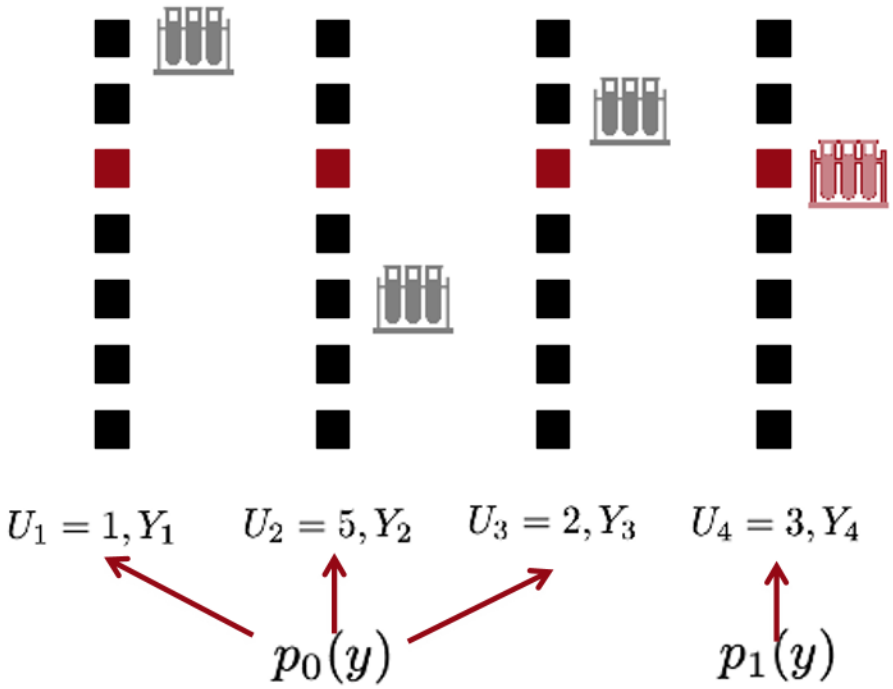
X = true system state

$$X = \begin{cases} 0 & \text{if no anomaly} \\ j & \text{if component } j \text{ anomalous} \end{cases}$$

$$X \in \{0, 1, \ldots, M\}$$

$$M = 7, X = 3$$

System Evolution



Person u

Observation y

conditional density

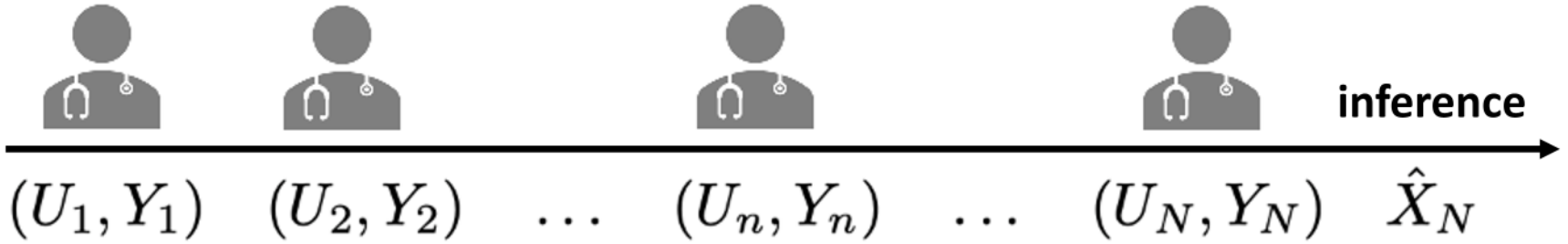
We assume these are known

We will need to learn these

Goals

USC Viterbi

- Experiment Selection Strategy:



$$U_n \sim g_n(I_n)$$

experiment choice – which person to test?

- Inference Strategy: decide **infected** or **not infected**

binary valued inference $\hat{X}_N \sim f(I_{N+1})$

not infected $X = 0$

infected $X \neq 0$

Max-min KL-Divergence

USC Viterbi

- Define

α, β distributions

$$D^* \doteq \max_{\alpha \in \Delta\mathcal{U}} \min_{j \in \mathcal{U}} \sum_{u \in \mathcal{U}} \alpha(u) D_j^u$$

argmax: α^*

$$= \min_{\beta \in \Delta\mathcal{U}} \max_{u \in \mathcal{U}} \sum_{j \in \mathcal{U}} \beta(j) D_j^u$$

argmin: β^*

- Lemma: we can compute D*

$$D^* = \left(\sum_{u \in \mathcal{U}} \frac{1}{D_u^u} \right)^{-1}$$

Non-asymptotic Bounds - Symmetric

USC Viterbi

- **Theorem:**

Strong converse: from decomposition and
strong converse in Polyanskiy, Poor and Verdu Trans IT 2010

$$-\log \phi_N^* \leq \text{INV}_N\left(\epsilon_N + \frac{\epsilon_N}{\eta}\right) + \log \frac{\eta}{\epsilon_N}$$

$$-\log \phi_N^* \geq \text{INV}_N\left(\epsilon_N - \frac{\epsilon_N}{\eta}\right) - O\left(\log \frac{\eta}{\epsilon_N}\right)$$

function of D*

- **Bounds enable the design of strategies**

INV_N: quantile function of $\bar{Z}_N + D(\beta^* || \tilde{\rho}_1)$

Comparisons

USC Viterbi

- Open-loop randomized (**OPE**): asymptotically optimal

  randomly select component from distribution α^*

  uniform in symmetric case

- Deterministic adaptive (**DAS**): also asymptotically optimal

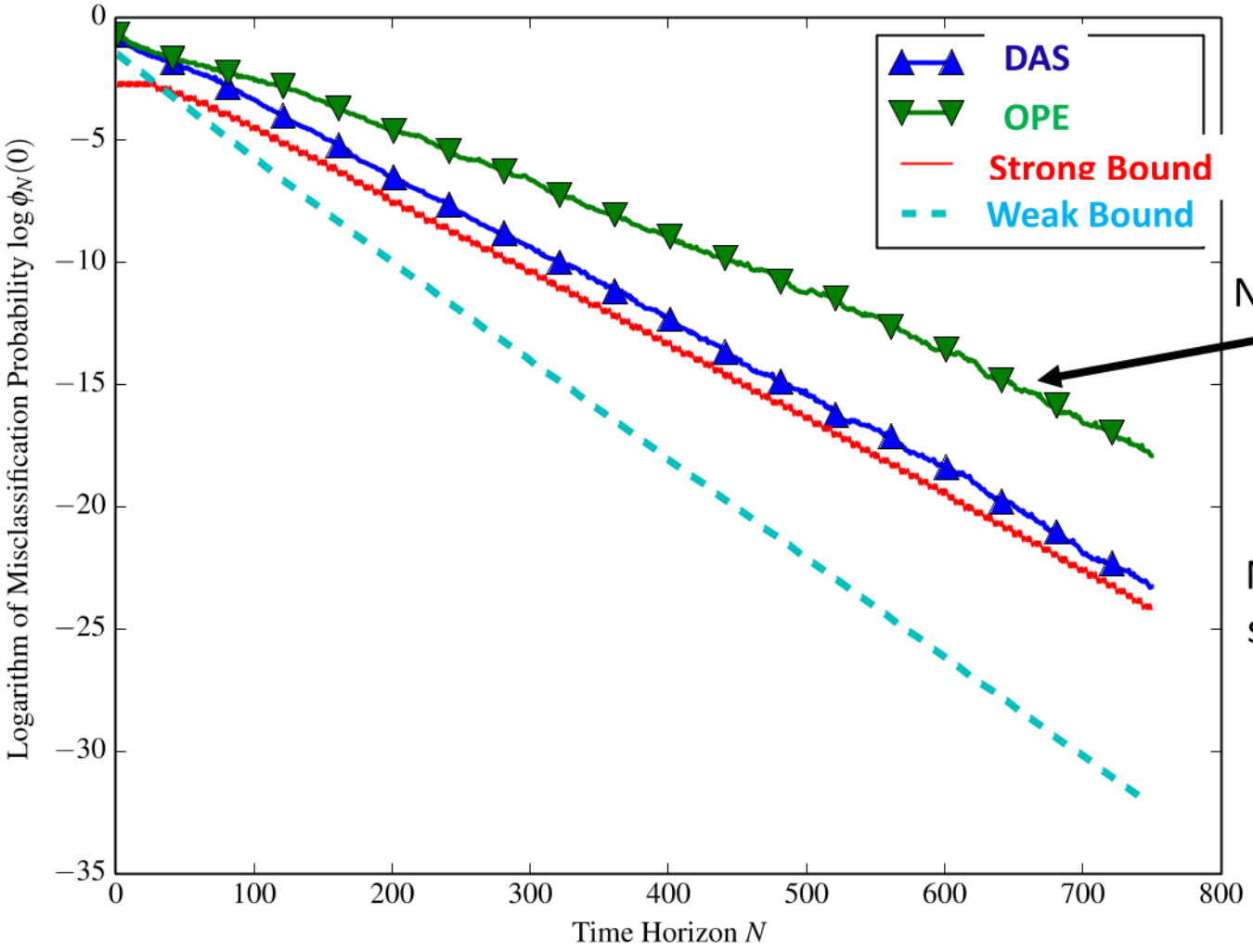
  at each time n, select the component j
  that minimizes $Z_{n-1}(j) - \log \tilde{\rho}_1(j)$

  function of previous observations and experiment choices

- Example setting: two-people and binary observations

Numerical Results

USC Viterbi

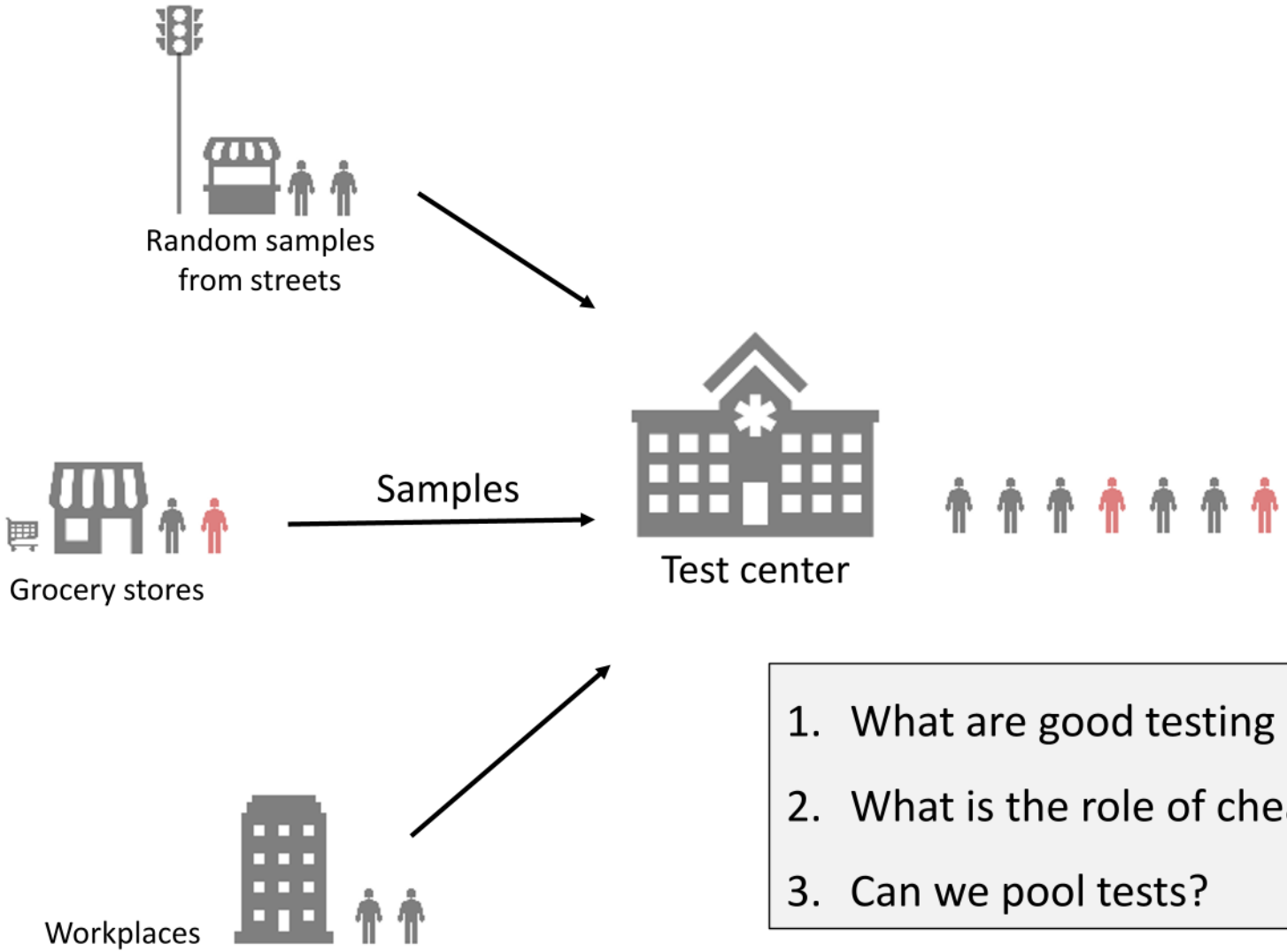


Open-loop:
Not second-order optimal

Need to be adaptive for second-order optimality

Practical SARS-CoV-2 Testing

USC Viterbi

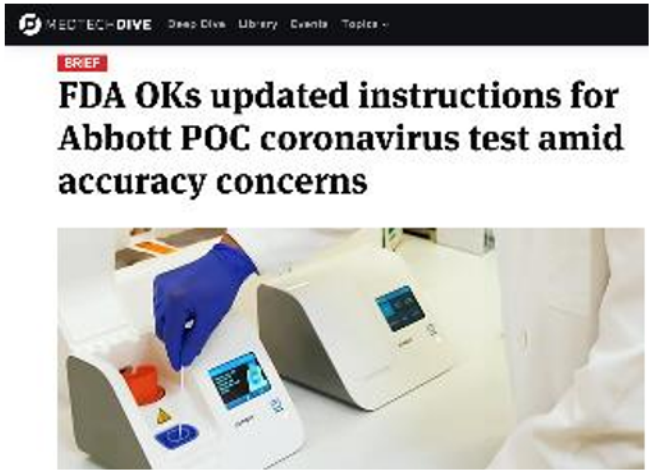


1. What are good testing strategies?
2. What is the role of cheap tests?
3. Can we pool tests?

Practical SARS-CoV-2 Testing

USC Viterbi

Random samples from streets



Grocery stores

Samples

Test center

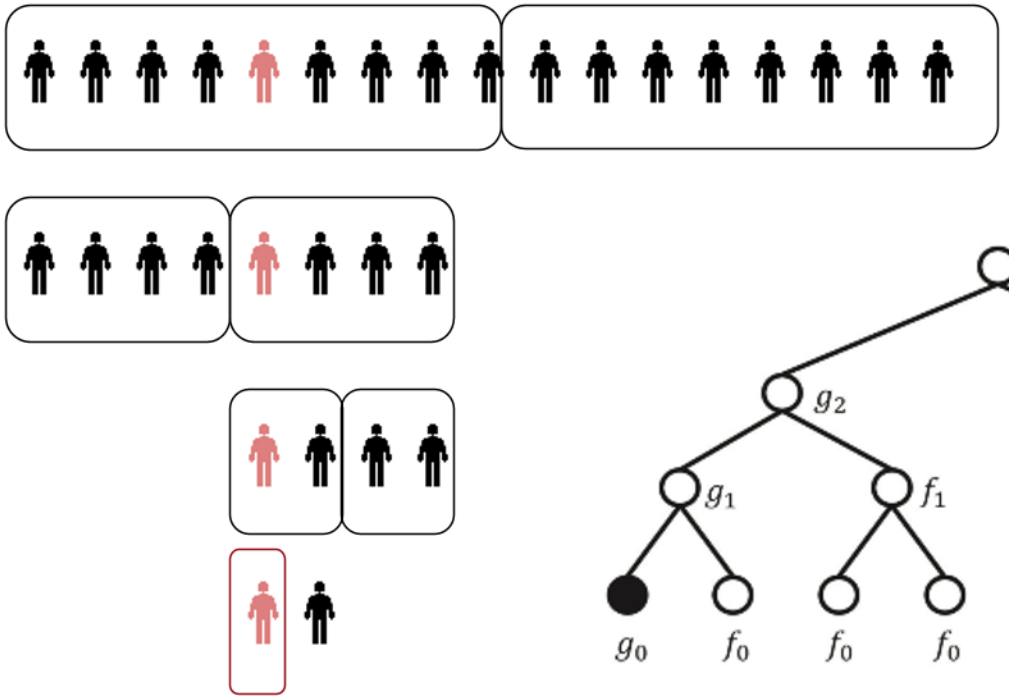
Workplaces

1. What are good testing strategies?
2. What is the role of cheap tests?
3. Can we pool tests?

Group Testing – pooling samples

USC Viterbi

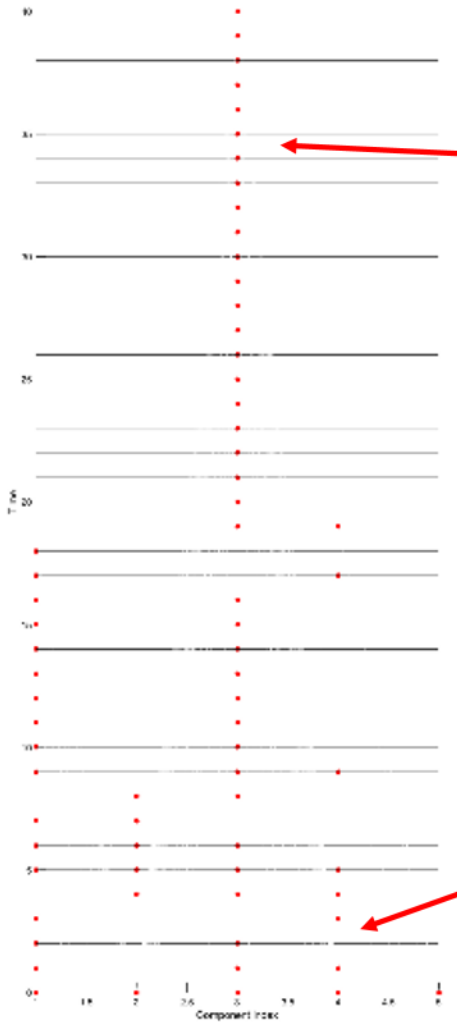
- Used in WW2 to test soldiers for syphilis
  - R. Dorfman, "The Detection of Defective Members of Large Populations," The Annals of Mathematical Statistics, 1943.



- N tests → log (N) tests

Prior Belief: At Most One Anomaly

USC Viterbi

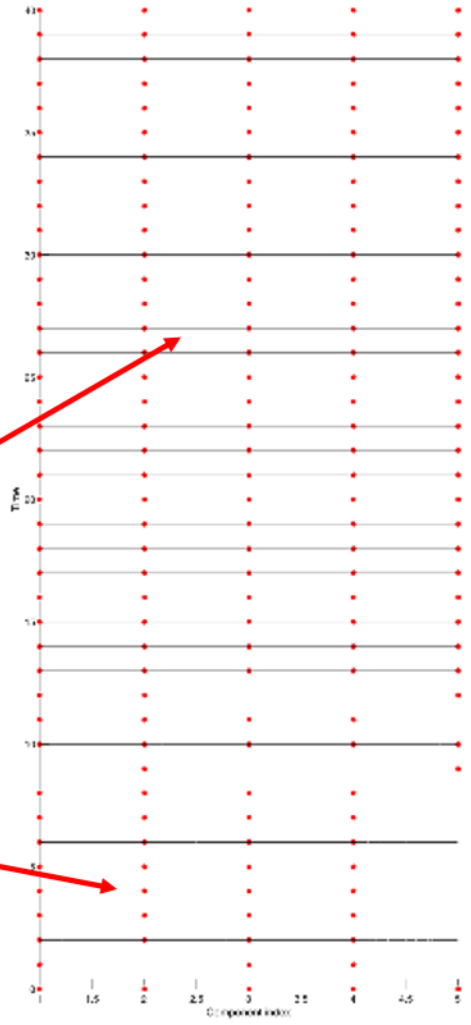


Single anomaly at index 3

Eventually anomaly is localized and sampled for confirmation

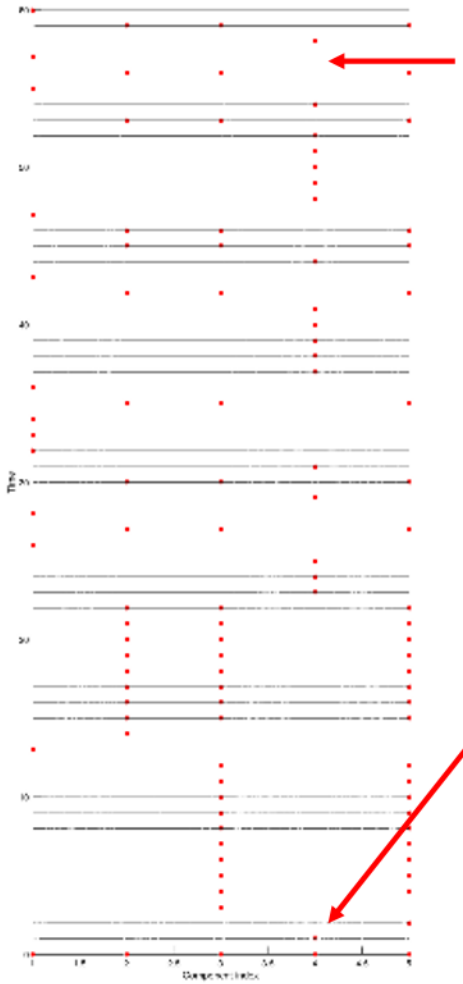
All components grouped together to confirm there is no anomaly

Initial groups aimed at searching



No anomaly

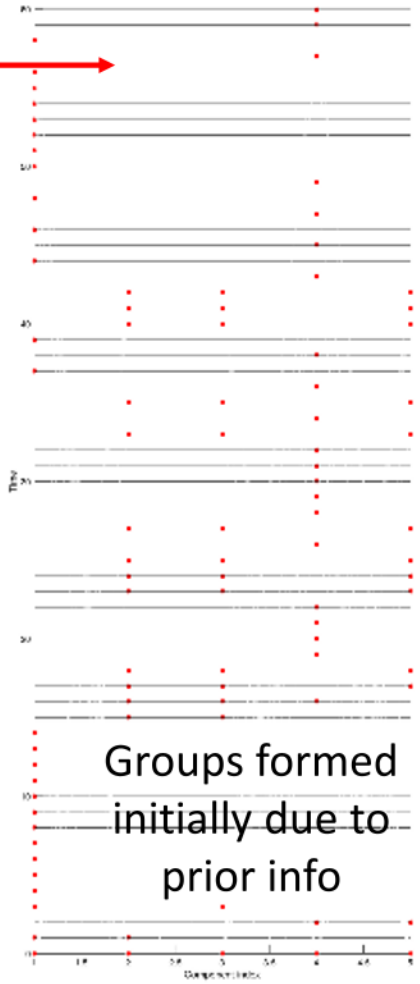
Two Anomalies



Confirmation: normal components are pooled and anomalies are tested individually

Initial tests are individual - known to be optimal with uniform prior - but groups are formed as belief improves

Anomalies at 1 and 4

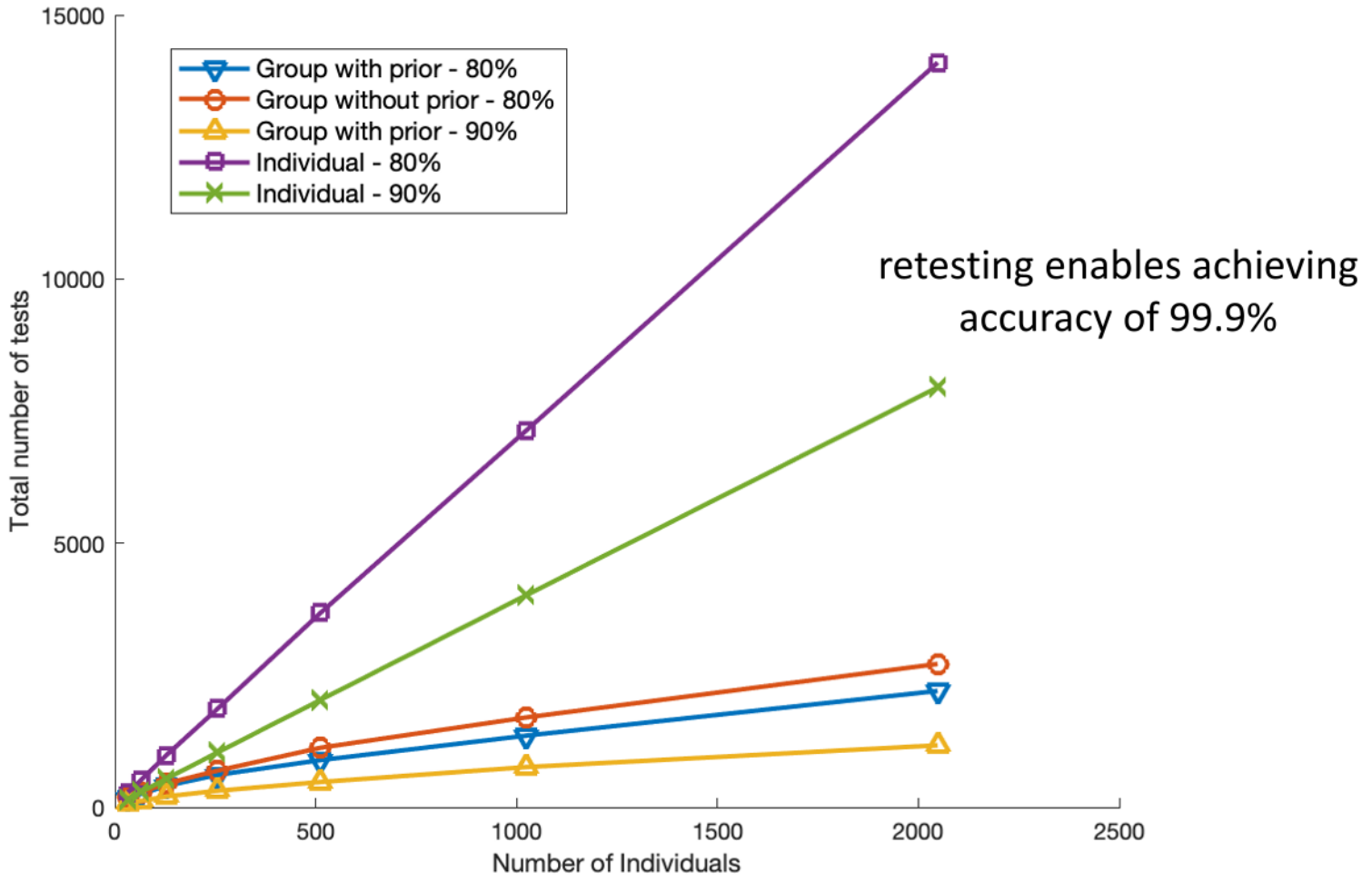


Prior: Any number anomalous

Prior: At most two anomalous

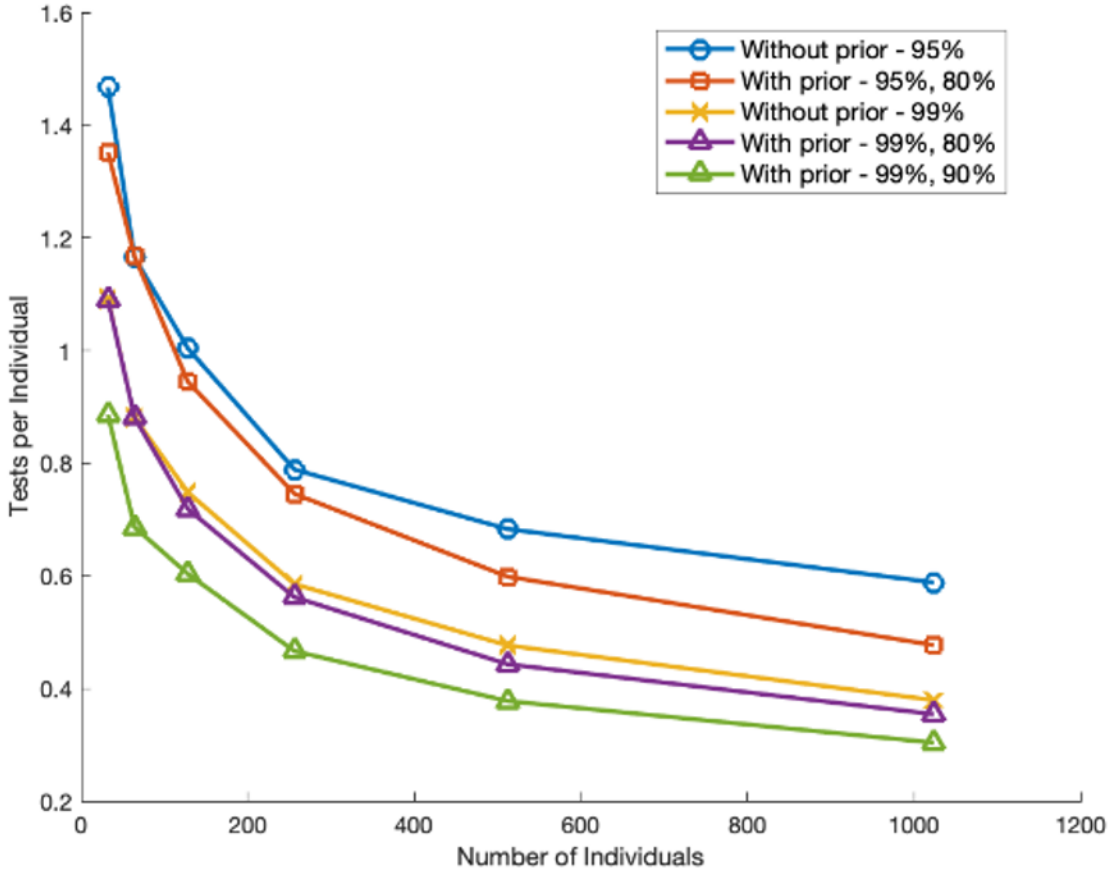
A single type of test

USC Viterbi



Fully-adaptive Tests

USC Viterbi



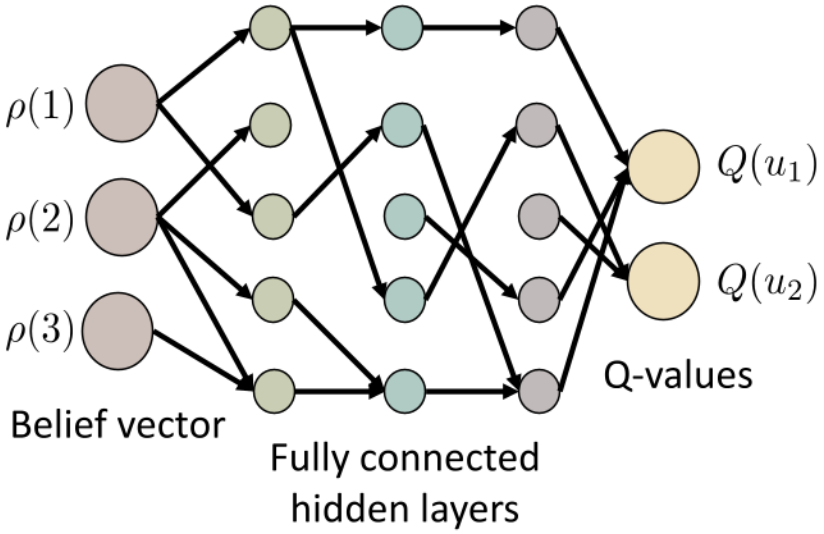
- Perform a cheap test first on each individual – we consider tests with 80% and 90% accuracy
- Use the prior for group testing subsequently
- Can reduce number of group tests by 20%

- Performing cheap tests first better when the cost of cheap test is about 10-15 times smaller

Fully adaptive tests can take a lot of time – need to parallelize

Challenges

- Optimal test design is computationally expensive
- We can exploit machine learning/neural network tools to compute optimal solution



- Have to do this carefully
  - recursive neural networks did not not work
  - Need the output of experiment sequences
- Exploit structural properties of optimal solution to design NN

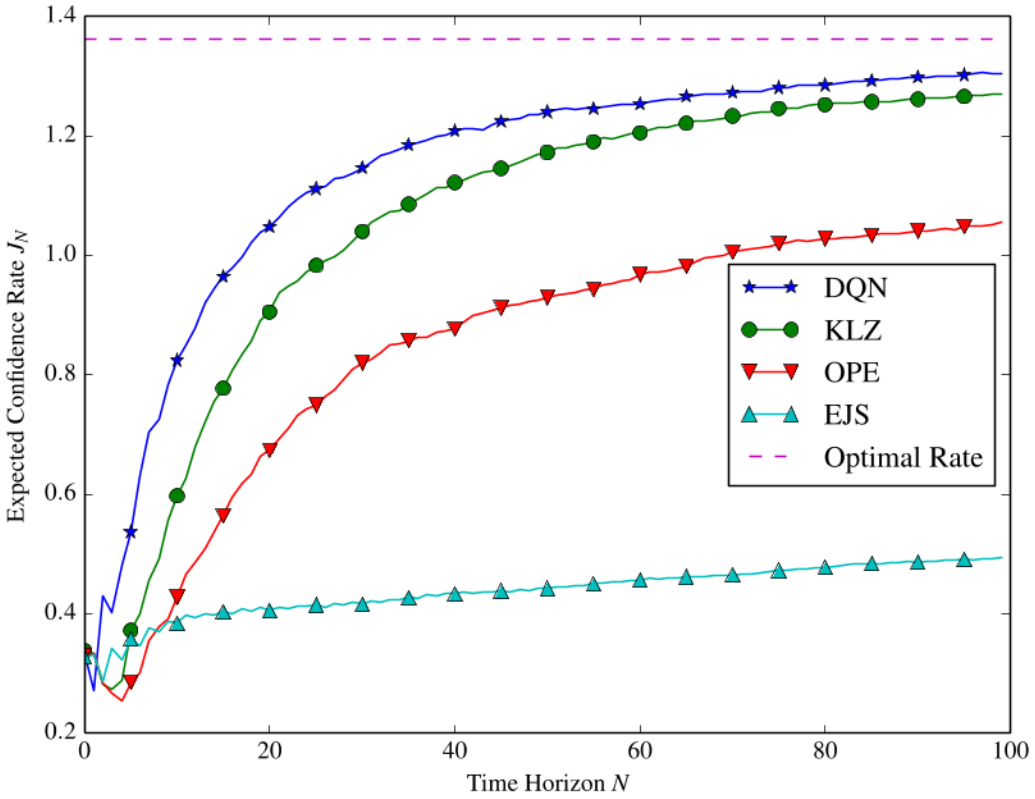
Deep Q Network

USC Viterbi

- Evolution of expected confidence under hypothesis h_0 over time

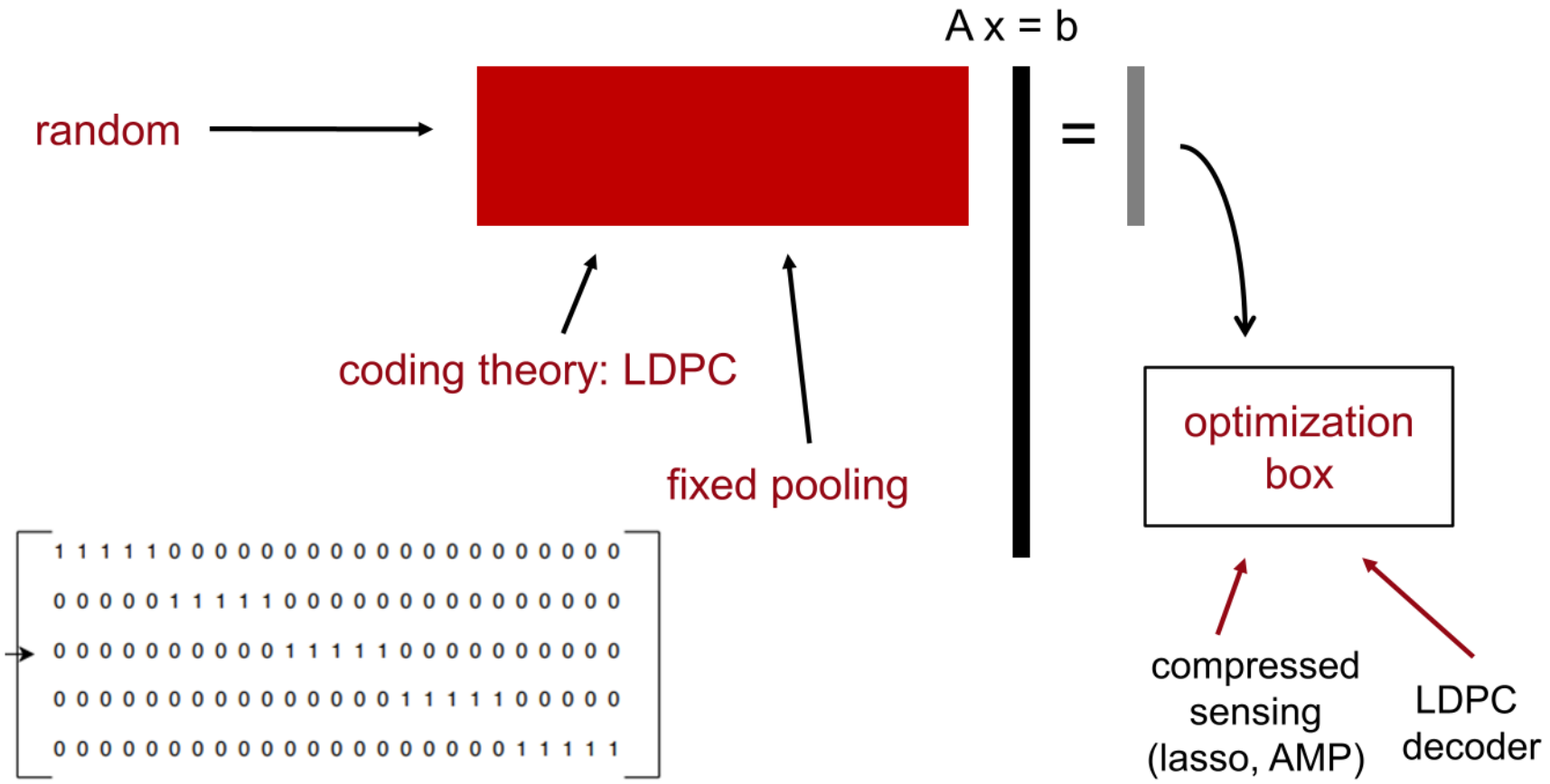
DQN learns the best policy

- **DAS** close to optimal rate
- **OPE** asymptotically optimal but very slow convergence
- **EJS** not optimal



Other Strategies

USC Viterbi

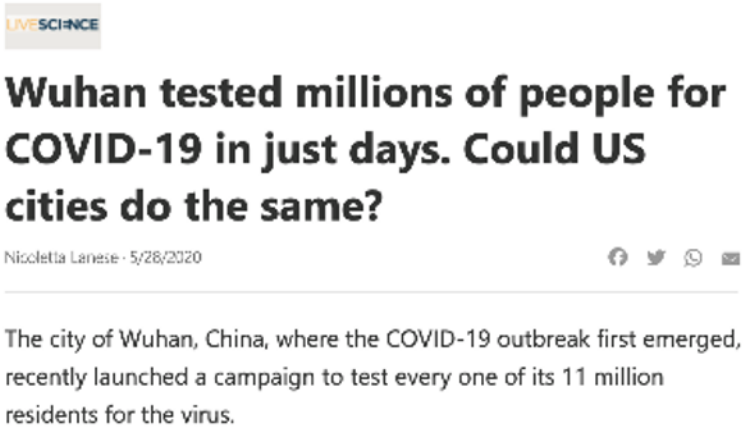


NON-ADAPTIVE, but can be parallelized

Contrasts

USC Viterbi

- Our method is data adaptive
  - More challenging to parallelize
- The test matters
  - Serological tests are blood based – easy to pool?
- Gold standard: PCR (polymerase chain reaction)
  - For SARS-CoV-2 test on RNA
  - Can parallelize

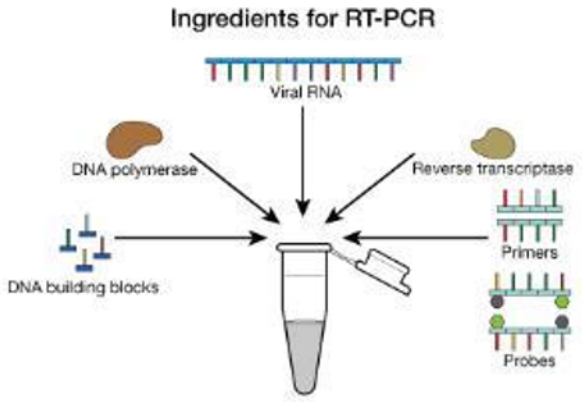


LIVESCIENCE

Wuhan tested millions of people for COVID-19 in just days. Could US cities do the same?

Nicoletta Lanese · 5/28/2020

The city of Wuhan, China, where the COVID-19 outbreak first emerged, recently launched a campaign to test every one of its 11 million residents for the virus.



Mayo Clinic explainer

Contrasts

USC Viterbi

- Our method is data adaptive
  - More challenging to parallelize
- The test matters
  - Serological tests are blood based – easy to pool?
- Gold standard: PCR (polymerase chain reaction)
  - For SARS-CoV-2 test on RNA
  - Can parallelize


UCSF

Testing the Tests: COVID-19 Antibody Assays Scrutinized for Accuracy by UCSF, UC Berkeley Researchers


NEWS | CORONAVIRUS (COVID-19) | JUNE 10, 2020

COVID-19 Genetic PCR Tests Give False Negative Results if Used Too Early

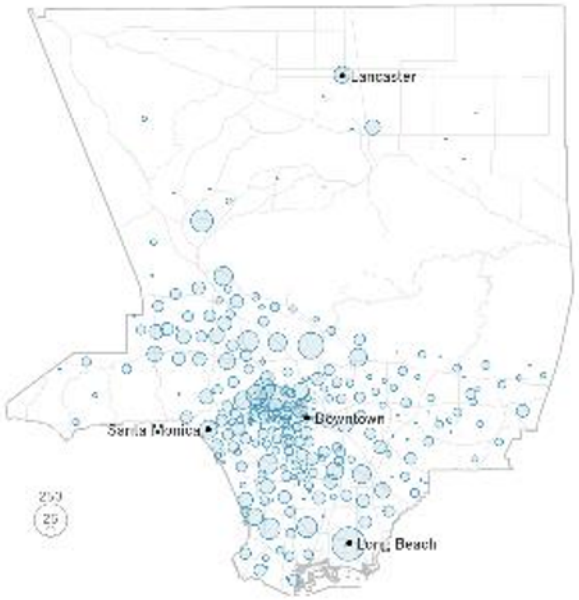
A new study confirms what many suspected, that PCR testing even 8 days after infection shows 20 percent false positives

June 10, 2020 — In a new study, Johns Hopkins researcher
testing people for SARS-CoV-2 (COVID-19) too early in the course of infection is likely to result in a false negative test, even though they may eventually test positive for the virus.[1] This is important to understand since many hospitals are using these COVID tests to screen patients before imaging exams, diagnostic testing or procedures.

Conclusions

USC Viterbi

- Optimized solutions for a finite number of observations/tests
  - Not asymptotics as in traditional methods
- We can design for both the exploration and the exploitation phases
- We can accommodate different kinds of information
  - Prior medical history, outcomes of other measurements (temperature, symptoms)
  - We can accommodate different kinds of SARS-CoV-2 tests, each with different efficacies
- Optimal testing for hot spots?
- Challenges
  - Complexity
  - Unknown onset
  - Parallelization



References

USC Viterbi

1. Zhu, Junan et al. “Noisy Pooled PCR for Virus Testing.” *bioRxiv* (2020)
2. Aldridge, Matthew. “Conservative two-stage group testing.” *ArXiv* abs/2005.06617 (2020)
3. Mutesa, Leon et al. “A strategy for finding people infected with SARS-CoV-2: optimizing pooled testing at low prevalence.” *medRxiv* (2020)
4. Ghosh, Sabyasachi et al. “A Compressed Sensing Approach to Group-testing for COVID-19 Detection.” *arXiv: Quantitative Methods* (2020)
5. Aldridge, Matthew. “Rates of Adaptive Group Testing in the Linear Regime.” *2019 IEEE International Symposium on Information Theory (ISIT)* (2019)
6. Kartik, Dhruva, Ashutosh Nayyar, and Urbashi Mitra. "Fixed-horizon Active Hypothesis Testing." *arXiv preprint arXiv:1911.06912* (2019). (Under review in TAC)
7. Kartik, Dhruva, Ashutosh Nayyar, and Urbashi Mitra. "Testing for Anomalies: Active Strategies and Non-asymptotic Analysis." *arXiv preprint arXiv:2005.07696* (2020).(Accepted in ISIT 2020)
8. Narayanan, Krishna R., Anoosheh Heidarzadeh, and Ramanan Laxminarayan. "On Accelerated Testing for COVID-19 Using Group Testing." *arXiv preprint arXiv:2004.04785* (2020).
9. Aldridge, Matthew, Oliver Johnson, and Jonathan Scarlett. "Group testing: an information theory perspective." *arXiv preprint arXiv:1902.06002* (2019).