Making Decisions with Shuffled Bits

Stefano Marano[(a)] Peter Willett[(b)]

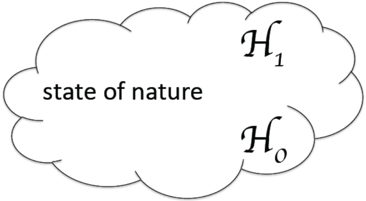
[(a)]DIEM, Univ. of Salerno, Fisciano (SA), Italy, marano@unisa.it

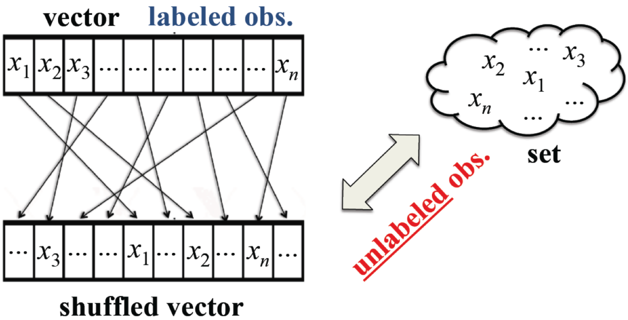
[(b)]ECE Dept., Univ. of Connecticut, Storrs, CT, USA, peter.willett@uconn.edu

2019 IEEE Intern. Conf. on Acoustics, Speech and Signal Processing

12 – 17 May, 2019 • Brighton, UK

Question Raised





With labeled obs.

$$\log \frac{p_1(x_1)\dots p_n(x_n)}{q_1(x_1)\dots q_n(x_n)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma$$

With unlabeled obs.

$p_i(\cdot)\, q_i(\cdot)$ known, but

… who goes with whom?

Application Areas & Motivation

- **Unlabeled SP:** Credit to [1] for initiating the field of unlabeled signal processing

- **Applications in:** Spoofing attacks to wireless ad-hoc nets or smart grids [2, 3]; big data scenarios (stripping time/space labels can be attractive [4]); image processing [5]; genome research [6]; archaeology [7]; communication over permutation channels [8]; molecular communications [9]. Further can be found in: [10, 11, 12, 13].

- **Motivational example** from Social Sensing:
  - Following the initiation of an **event** meant to be **covert**, users take consequent actions (visit specific webpages, post comments, contact friends, . . .).
  - Users are ***profiled***: A network analyzer knows the probability that each user takes an action as consequence of each hidden event

    $\Rightarrow$ *event can be therefrom inferred*
  - What if **users' actions** are **anonymized**? Can the covert event be still inferred by users' profiles? And how powerful is such a labeled-unaware network analyzer?

- **At more theoretical level:** In a detection problem, how much **information** is contained in the observation **values**, and how much in their **labels**?

Formalization

Labeled observations

- (Labeled) Binary Observations: $\mathbf{X}^n \sim \prod_{i=1}^{n} r_i^{x_i}(1-r_i)^{1-x_i}$, $X_i \in \{0,1\}$
- Statistical Test: $\begin{array}{l} \mathcal{H}_1 : r_i = p_i = \mathbb{P}_1(X_i = 1), \\ \mathcal{H}_0 : r_i = q_i = \mathbb{P}_0(X_i = 1), \end{array}$ $\quad i = 1, 2, \ldots, n$
- Solution: LLR $\quad \sum_{i=1}^{n} x_i \log \frac{p_i}{q_i} + (1-x_i) \log \frac{1-p_i}{1-q_i} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma$

Unlabeled observations

- Unlabeled Binary Observations: $\mathbf{X}^n \sim \prod_{i=1}^{n} r_i^{x_{\pi(i)}}(1-r_i)^{1-x_{\pi(i)}}$, π unknown
- What test? GLRT is a possibility: replace π by its ML estimate $\hat{\pi}$

$$\sum_{i=1}^{n} \left[x_{\hat{\pi}_1(i)} \log p_i + (1 - x_{\hat{\pi}_1(i)}) \log(1-p_i) \right] - \sum_{i=1}^{n} \left[x_{\hat{\pi}_0(i)} \log q_i + (1 - x_{\hat{\pi}_0(i)}) \log(1-q_i) \right]$$

GLRT (1/2)

- ML estimate under $\mathcal{H}_1$: $\hat{\pi}_1 = \arg\max_{\pi} \log \prod_{i=1}^{n} p_i^{x_{\pi(i)}} (1-p_i)^{1-x_{\pi(i)}}$

  $\Leftrightarrow$ Find the best path over the trellis

$$\begin{pmatrix} \log p_1 & \log p_2 & \log p_3 & \dots & \log p_n \\ \log(1-p_1) & \log(1-p_2) & \log(1-p_3) & \dots & \log(1-p_n) \end{pmatrix}$$

- ML estimate under $\mathcal{H}_0$: $\hat{\pi}_0 = \arg\max_{\pi} \log \prod_{i=1}^{n} q_i^{x_{\pi(i)}} (1-q_i)^{1-x_{\pi(i)}}$

  $\Leftrightarrow$ Find the best path over the trellis

$$\begin{pmatrix} \log q_1 & \log q_2 & \log q_3 & \dots & \log q_n \\ \log(1-q_1) & \log(1-q_2) & \log(1-q_3) & \dots & \log(1-q_n) \end{pmatrix}$$

GLRT (2/2)

With arbitrary alphabets (known facts)

- Algorithmic approach via *assignment problem*
- Hungarian (Munkres), JVC and auction algorithms have been advocated
- **No closed-form solution**; **complexity** is an issue

With binary alphabets

Result 1. The GLRT statistic is given by

$$\mathcal{S}_{\mathrm{GLRT}} = \sum_{i=1}^{k_{\mathbf{x}}} \log \frac{p_{(i)}}{q_{(i)}} + \sum_{i=k_{\mathbf{x}}+1}^{n} \log \frac{1-p_{(i)}}{1-q_{(i)}}$$

$k_{\mathbf{x}}$ = number of ones

$p_{(i)}$ = i-th largest element of $(p_1, p_2, \ldots, p_n)$

Detectors A and B

With arbitrary alphabets (known facts)

- Two greedy algorithms have been proposed as surrogates for the GLRT
  - *Detector A* sequentially matches observations to "most convenient" distribution: *Greedy search of best path over the trellis*
  - *Detector B* first finds the best sequence of distributions, then sequentially adapts the sequence to observations: *Greedy adaptation of best sequence to observations*
- Complexity upper bounded by $\mathcal{O}(n^2)$
- Relative merits & actual complexity remain open issues

With binary alphabets

Result 2. Detectors A and B coincide, and both are equivalent to GLRT

ULR Detector (1/2)

With arbitrary alphabets ($|\mathcal{X}| > 2$):

- $\widetilde{\mathbf{X}}^n = (\widetilde{X}_1, \ldots, \widetilde{X}_n)$, $\widetilde{X}_i$ *iid* $\sim \bar{p} = \frac{1}{n}\sum_{i=1}^n p_i$ or $\widetilde{X}_i$ *iid* $\sim \bar{q} = \frac{1}{n}\sum_{i=1}^n q_i$
- $t_{\widetilde{\mathbf{x}}^n}$ **type** of *iid* observations
  - SLLN $t_{\widetilde{\mathbf{X}}^n}(x) \to \bar{p}(x)$ ae under $\mathcal{H}_1$, $t_{\widetilde{\mathbf{X}}^n}(x) \to \bar{q}(x)$ ae under $\mathcal{H}_0$
- $t_{\mathbf{x}^n}$ **type** of observations
  - $\mathrm{VAR}_{1,0}[\mathcal{I}(X_i = x)] = r_i(x)(1 - r_i(x))$, $\sum_{i=1}^{\infty} \mathrm{VAR}_{1,0}[\mathcal{I}(X_i = x)]/i^2 \le \frac{\pi^2}{24} < \infty$
  - $\Rightarrow \frac{1}{n}\sum_{i=1}^{n} \mathcal{I}(X_i = x) - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_{1,0}[\mathcal{I}(X_i = x)] \to 0$
  - $\Rightarrow \forall \epsilon > 0$ and n sufficiently large $|t_{\mathbf{X}^n}(x) - t_{\widetilde{\mathbf{X}}^n}(x)| < \epsilon$ ae

ULR Detector (2/2)

With arbitrary alphabets (known facts)

- Type vector $t_{\mathbf{x}}$ and type vector from *iid* obs. $t_{\widetilde{\mathbf{x}}}$, converge to the same constant vector
- LLR for *iid* obs.: $\sum_{x \in \mathcal{X}} t_{\widetilde{\mathbf{x}}}(x) \frac{\bar{p}(x)}{\bar{q}(x)}$ ULR [12]: $\sum_{x \in \mathcal{X}} t_{\mathbf{x}}(x) \frac{\bar{p}(x)}{\bar{q}(x)}$
- By simulations: nice performance in many cases (perhaps unexpectedly)
- (Analytical) Performance assessment is an **open issue**
- Relative merit wrt GLRT, Detector A, Detector B, **mostly unexplored**

With binary alphabets

Result 3. ULR reduces to a simple *counting* detector:

$$\mathcal{S}_{\mathrm{ULR}} = \mathrm{sign}(\bar{p} - \bar{q})\ k_{\mathbf{x}}$$

Finite No. of Classes

Suppose:

$$p = (\underbrace{p_{c1}, \dots, p_{c1}}_{n_1}, \underbrace{p_{c2}, \dots, p_{c2}}_{n_2}, \dots\dots, \underbrace{p_{cm}, \dots, p_{cm}}_{n_m})$$

$$q = (\underbrace{q_{c1}, \dots, q_{c1}}_{n_1}, \underbrace{q_{c2}, \dots, q_{c2}}_{n_2}, \dots\dots, \underbrace{q_{cm}, \dots, q_{cm}}_{n_m})$$

Detector for shuffled bits:

- $k_{\mathbf{X}} = \sum_{i=1}^{n_1} X_i + \sum_{i=n_1+1}^{n_1+n_2} X_i + \cdots + \sum_{\sum_{k=1}^{m-1} n_k+1}^{n} X_i$. By CLT arguments (n_ℓ large)

Result 4

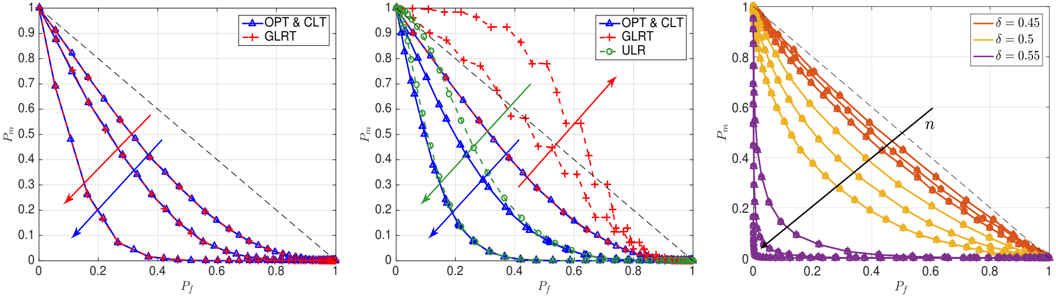
$$\mathcal{S}_{\text{CLT}} = \left(\frac{k_{\mathbf{x}} - n\bar{q}}{\sigma_0}\right)^2 - \left(\frac{k_{\mathbf{x}} - n\bar{p}}{\sigma_1}\right)^2$$

where $\sigma_1^2 = \sum_{\ell=1}^{m} n_\ell \, p_{c\ell} \, (1 - p_{c\ell})$

– boils down to ULR for $\sigma_1 = \sigma_0$

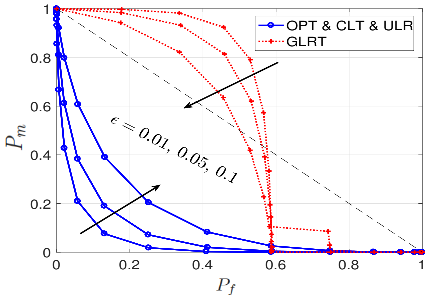
– **works beyond the m-class setting**

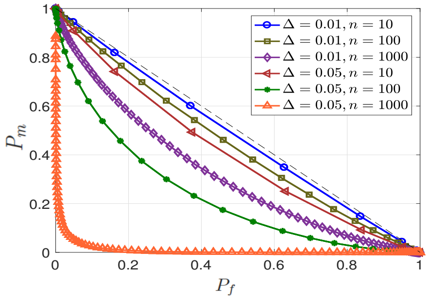
Simulation Results (1/3)



- *Left:* $m = 2$ classes, $n_1 = n_2 = 100$. $q_{c1} = q_{c2} = .5$. Following the arrows: $(p_{c1} = .9, p_{c2} = .1)$, $(p_{c1} = .95, p_{c2} = .05)$, and $(p_{c1} = .99, p_{c2} = .01)$.
- *Middle:* Same as in *Left*, except: $(p_{c1} = .9, p_{c2} = .1)$, $(p_{c1} = .95, p_{c2} = .1)$, $(p_{c1} = .99, p_{c2} = .1)$.
- *Right:* $m = 10$ classes, each with n/m entries, and $n = 50, 100, 200$. q_{ci}, generated uniformly at random $\in (.45, .55)$, p_{ci}, uniformly at random $\in (\delta, \delta + 0.1)$.

Simulation Results (2/3)





- *Left:* $m = 2$ classes, $n_1 = n_2 = 10$. $q_{c1} = q_{c1} = .5$, $p_{c1} = 1 - \epsilon$, $p_{c2} = 1/2 - \epsilon$.
- *Right:* $(q_1, \ldots, q_n)$ grows linearly from $q_1 = 0.3$ to $q_n = 0.7$, $(p_1, \ldots, p_n)$ grows linearly from $p_1 = 0.3 + \Delta$ to $p_n = 0.7 + \Delta$, where $\Delta = 0.01, 0.05$, and $n = 10, 10^2, 10^3$.

Simulation Results (3/3)

- **Two classes (optimum easily computable):**
  - $\mathcal{H}_0$: balan. *iid* vs $\mathcal{H}_1$: half obs. $\mathbb{P}_1(X_i = 1) = 1 - \epsilon$, half $\mathbb{P}_1(X_i = 1) = \epsilon$
    CLT same as OPT, **GLRT quite close to OPT**, **ULR useless**
  - $\mathcal{H}_0$: balan. *iid* vs $\mathcal{H}_1$: half obs. $\mathbb{P}_1(X_i = 1) = \epsilon$, half $\mathbb{P}_1(X_i = 1) = \frac{1}{2} - \epsilon$
    CLT & ULR same as OPT, **GLRT useless**

- **General considerations:**
  - **GLRT** should be used only after checking its **unbiasedness**
  - **ULR** is expected to work well, **unless** $|\bar{p} - \bar{q}|$ **is too small**
  - **CLT** is recommended also in **challenging scenarios**

- **General trends:**
  - Performance improves with n and with distribution "distance" ...
  - ... *in primis*: **how** $\bar{p}$ **is far from** $\bar{q}$, *in secundis* **how** σ_1 **is far from** σ_0
  - $\bar{p} = \bar{q} \quad \Rightarrow \quad P_m, P_d$ scale **sub-exponentially** with $n \to \infty$
  - $\bar{p} \approx \bar{q} \quad \Rightarrow \quad$ Performance **only depends** on: $|\bar{p} - \bar{q}|$, σ_1, σ_0

Conclusions (1/2)

GLRT with $|\mathcal{X}| > 2$

- GLRT boils down to solving an assignment problem
- There are cases in which GLRT is useless
  - Non consistent: the search space grows more than exponentially fast with n
- Detectors A and B: Relative merits? Performance? Relation to GLRT?

GLRT with Binary Observations

- Simple closed-form expression (performance assessment possible)
- Modest computational cost
- Same of Detectors A and B
- There exist detection problems in which GLRT is biased (more in [13])

Conclusions (2/2)

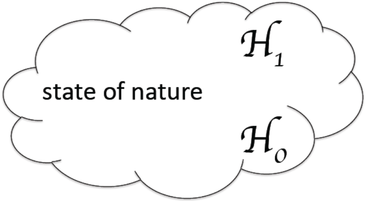
ULR

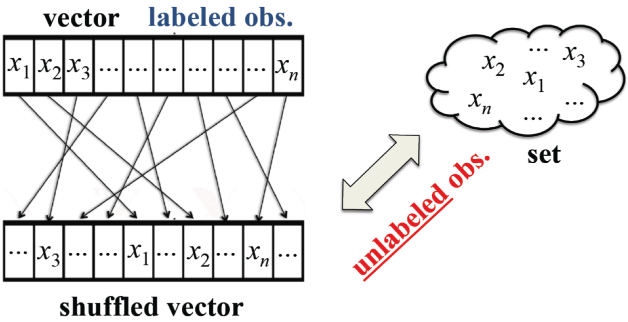
- Computationally very efficient
- Works when $|\bar{p} - \bar{q}|$ is not too small
- With $|\mathcal{X}| > 2$: Performance assessment? Relative merits wrt other det.?
- With **binary observations**: boils down to a counting detector

CLT for m-class binary observations

- Good trade off between complexity/performance
- Exploits diversity in mean and in variance

Coming Soon . . . (1/2)





With labeled obs.

Fundamental limit:
error exponent $\Omega(\alpha)$

With unlabeled obs.

Fundamental limit?

Coming Soon ... (2/2)

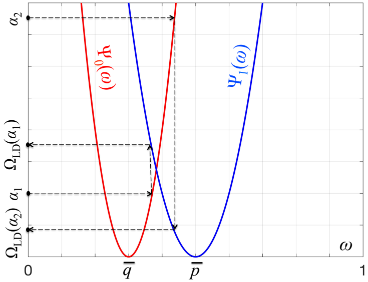
With arbitrary alphabets (known facts)

$\psi_1(\lambda) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \log \sum_{x\in\mathcal{X}} p_i(x) e^{\lambda(x)}$

$\Psi_1(\omega) = \mathrm{LT}[\psi_1(\lambda)] = \sup_{\lambda\in\mathfrak{R}^{|\mathcal{X}|-1}} \{\sum_{x\in\mathcal{X}'} \lambda(x)\omega(x) - \psi_1(\lambda)\}$

$$\Omega(\alpha) = \inf_{\omega\in\mathcal{P}(\mathcal{X}):\, \Psi_0(\omega)<\alpha} \Psi_1(\omega)$$

With binary alphabets (and low-detectability regime)



$$\Omega(\alpha) \approx \frac{\left(\left[|\bar{p}-\bar{q}| - \sqrt{2\bar{\sigma}_0^2\,\alpha}\,\right]^+\right)^2}{2\bar{\sigma}_1^2}$$

with $\bar{\sigma}_1^2 = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} p_i(1-p_i)$

Essential Bibliography

[1] J. Unnikrishnan, S. Haghighatshoar, and M. Vetterli, "Unlabeled sensing with random linear measurements," *IEEE Transactions on Information Theory*, vol. 64, no. 5, pp. 3237–3253, May 2018.

[2] P. Pradhan, K. Nagananda, P. Venkitasubramaniam, S. Kishore, and R. S. Blum, "GPS spoofing attack characterization and detection in smart grids," *2016 CNS*, Oct 2016, pp. 391–395.

[3] Z. Zhang, S. Gong, A. D. Dimitrovski, and H. Li, "Time synchronization attack in smart grid: Impact and analysis," *IEEE Transactions on Smart Grid*, vol. 4, no. 1, pp. 87–98, March 2013.

[4] L. Keller, M. J. Siavoshani, C. Fragouli, K. Argyraki, and S. Diggavi, "Identity aware sensor networks," *Proc. of the 26th IEEE (INFOCOM 2009)*, Rio De Janeiro, Brazil, April, 19-25 2009, pp. 2177–2185.

[5] P. David, D. Dementhon, R. Duraiswami, and H. Samet., "SoftPOSIT: Simultaneous pose and correspondence determination," *International Journal of Computer Vision*, vol. 59, no. 3, pp. 259–284, 2004.

[6] X. Huang and A. Madan, "CAP3: A DNA sequence assembly program," *Genome Research*, vol. 9, pp. 868–877, 1999.

[7] W. S. Robinson, "A method for chronologically ordering archaeological deposits," *American Antiquity*, pp. 293–301, 1951.

[8] L. J. Schulman and D. Zuckerman, "Asymptotically good codes correcting insertions, deletions, and transpositions," *IEEE Transactions on Information Theory*, vol. 45, no. 7, pp. 2552–2557, 1999.

[9] T. Nakano, A. W. Eckford, and T. Haraguchi, *Molecular Communication*. UK: Cambridge University Press, 2013.

[10] S. Marano, V. Matta, P. Willett, P. Braca, and R. Blum, "Hypothesis testing in the presence of Maxwell's daemon: Signal detection by unlabeled observations," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2017)*, New Orleans, LA, USA, 5-9 Mar. 2017.

[11] G. Wang, J. Zhu, R. S. Blum, P. Willett, S. Marano, V. Matta, and P. Braca, "Signal amplitude estimation and detection from unlabeled binary quantized samples," *IEEE Transactions on Signal Processing*, vol. 66, no. 16, pp. 4291–4303, Aug. 2018.

[12] S. Marano and P. Willett, "Algorithms and fundamental limits for unlabeled detection using types," *IEEE Transactions on Signal Processing*, vol. 67, no. 8, pp. 2022–2035, Apr. 2019.

[13] S. Marano and P. Willett, "Making decisions by unlabeled bits," *to be submitted*