COMMUNICATION-CONSTRAINED INFERENCE AND THE ROLE OF SHARED RANDOMNESS

Clément Canonne (Stanford University)

June 13, 2019

With **Jayadev Acharya** (Cornell University) and **Himanshu Tyagi** (IISc Bangalore)

MOTIVATION: A TALE

Boaty McBoatface is starting its first mission today!
It's going to Antarctica to study global warming, not to play.

The world's oceans are changing, you see.
It's freezing down there, but not as cold as it used to be.

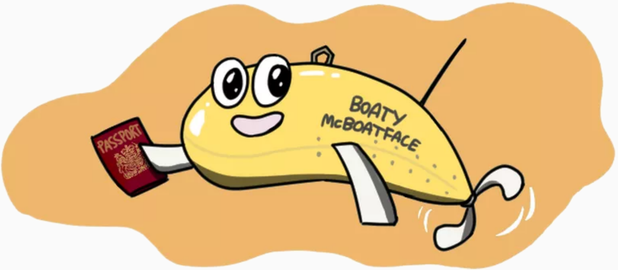


Illustration ©Dami Lee

Boaty's findings will be sent to scientists with care,
By way of a radio link, but with a certain flair.

Illustration ©Dami Lee

McBoatfaces are expensive

What is the most ship-efficient protocol to reliably test whether the distribution of temperatures matches the one on record?

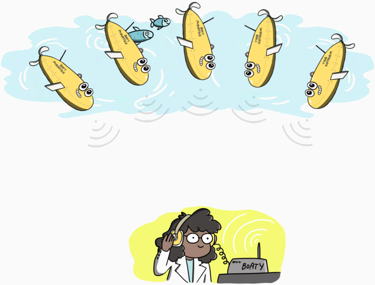


Illustration ©Dami Lee

DISTRIBUTED INFERENCE

THE SETTING: “SIMULTANEOUS COMMUNICATION PROTOCOL” (SMP)

the setting: "simultaneous communication protocol" (smp)

- an inference task $\mathcal{P}$ over k-ary distributions

THE SETTING: "SIMULTANEOUS COMMUNICATION PROTOCOL" (SMP)

- an inference task $\mathcal{P}$ over k-ary distributions
- an unknown k-ary distribution p

THE SETTING: "SIMULTANEOUS COMMUNICATION PROTOCOL" (SMP)

- an inference task $\mathcal{P}$ over k-ary distributions
- an unknown k-ary distribution p
- one centralized "referee" $\mathcal{R}$ who needs to solve $\mathcal{P}$ on p

the setting: "simultaneous communication protocol" (smp)

- an inference task $\mathcal{P}$ over k-ary distributions
- an unknown k-ary distribution p
- one centralized "referee" $\mathcal{R}$ who needs to solve $\mathcal{P}$ on p
- communication constraints represented by the set of "allowed channels" $\mathcal{W} = \{W\colon [k] \to \{0,1\}^\ell\}$
- n players, each choosing a channel $W \in \mathcal{W}$

THE SETTING: "SIMULTANEOUS COMMUNICATION PROTOCOL" (SMP)

- an inference task $\mathcal{P}$ over k-ary distributions
- an unknown k-ary distribution p
- one centralized "referee" $\mathcal{R}$ who needs to solve $\mathcal{P}$ on p
- communication constraints represented by the set of "allowed channels" $\mathcal{W} = \{W\colon [k] \to \{0,1\}^\ell\}$
- n players, each choosing a channel $W \in \mathcal{W}$
- each player independently gets one sample x from p and sends a message $y = W(x)$ to $\mathcal{R}$

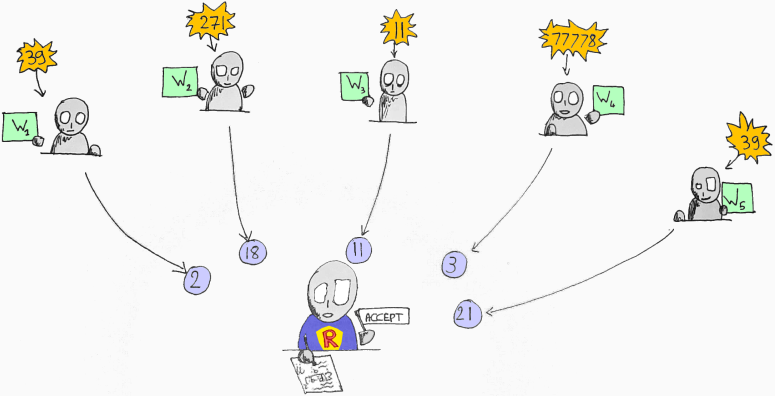
THE SETTING: "SIMULTANEOUS COMMUNICATION PROTOCOL" (SMP)

- an inference task $\mathcal{P}$ over k-ary distributions
- an unknown k-ary distribution p
- one centralized "referee" $\mathcal{R}$ who needs to solve $\mathcal{P}$ on p
- communication constraints represented by the set of "allowed channels" $\mathcal{W} = \{W : [k] \to \{0,1\}^\ell\}$
- n players, each choosing a channel $W \in \mathcal{W}$
- each player independently gets one sample x from p and sends a message $y = W(x)$ to $\mathcal{R}$

Question

> As a function of k, ℓ, and all relevant parameters of $\mathcal{P}$, what is the number of players n required?

THE SETTING, CONT'D



THE SETTING, CONT'D

THE SETTING, CONT'D

- if $\ell \geq \log_2 k$, trivial (no constraints)

THE SETTING, CONT'D

- if $\ell \geq \log_2 k$, trivial (no constraints)
- Inference tasks: density estimation, parameter estimation, functional estimation, hypothesis testing/property testing…

THE SETTING, CONT'D

- if $\ell \geq \log_2 k$, trivial (no constraints)
- Inference tasks: density estimation, parameter estimation, functional estimation, hypothesis testing/property testing...
- Different resources: public-coin, private-coin

Public-coin protocols players share a common random seed (e.g., broadcast by the server)
$\rightsquigarrow (W_1, \dots, W_n)$ jointly randomized

Private-coin protocols players have their own randomness only
$\rightsquigarrow (W_1, \dots, W_n)$ independent

Public-coin protocols players share a common random seed (e.g., broadcast by the server)
$\rightsquigarrow (W_1, \dots, W_n)$ jointly randomized

Private-coin protocols players have their own randomness only
$\rightsquigarrow (W_1, \dots, W_n)$ independent

In both cases, no communication between players.

ENOUGH WITH THE FANCY "P"... WHAT ARE WE TALKING ABOUT ANYWAY?

Focused on two specific fundamental* inference tasks:

Distribution Learning (estimation)

Must output: $\hat{p}$ s.t. $\ell_1(p, \hat{p}) \leq \varepsilon$

(and be correct on any p with probability at least 2/3)

ENOUGH WITH THE FANCY "P"... WHAT ARE WE TALKING ABOUT ANYWAY?

Focused on two specific fundamental* inference tasks:

Distribution Learning (estimation)

Must output: $\hat{p}$ s.t. $\ell_1(p, \hat{p}) \leq \varepsilon$

(and be correct on any p with probability at least 2/3)

Uniformity Testing (goodness-of-fit)

Must decide: $p = u_k$ (uniform), or $\ell_1(p, u_k) > \varepsilon$?

(and be correct on any p with probability at least 2/3)

* "If we can make it here, we can make it anywhere."

What is known without local constraints:

Task $\mathcal{P}$	n
Distribution learning	$\frac{k}{\varepsilon^2}$
Uniformity testing	$\frac{\sqrt{k}}{\varepsilon^2}$

What is known without local constraints:

Task $\mathcal{P}$	n
Distribution learning	$\frac{k}{\varepsilon^2}$
Uniformity testing	$\frac{\sqrt{k}}{\varepsilon^2}$

What happens with them?

What is known without local constraints:

Task $\mathcal{P}$	n
Distribution learning	$\frac{k}{\varepsilon^2}$
Uniformity testing	$\frac{\sqrt{k}}{\varepsilon^2}$

What happens with them? And does public randomness help then?

Our results with local constraints:

Task $\mathcal{P}$	n (private-coin)	n (public-coin)
Distribution learning	$\frac{k}{\varepsilon^2} \cdot \frac{k}{2^\ell}$	$\frac{k}{\varepsilon^2} \cdot \frac{k}{2^\ell}$
Uniformity testing	$\frac{\sqrt{k}}{\varepsilon^2} \cdot \frac{k}{2^\ell}$	$\frac{\sqrt{k}}{\varepsilon^2} \cdot \sqrt{\frac{k}{2^\ell}}$

PLAN FOR THE TALK

1. Private-Coin Swiss Army Knife: “Simulate-and-Infer”
2. Public-Coin Uniformity Testing: “Minimally Contracting Hashing”
3. Conclusion

“SIMULATE-AND-INFER”

ONE APPROACH TO SOLVE IT ALL

Key Observation

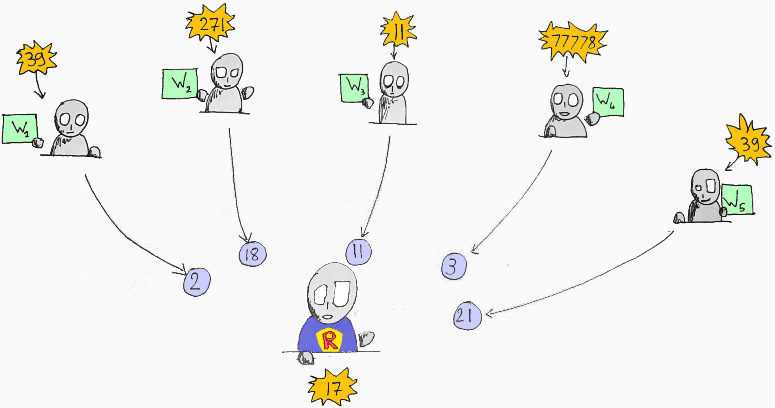
If the referee can simulate independent samples from p using the messages from the players, then it can do anything.

Key Observation

If the referee can simulate independent samples from p using the messages from the players, then it can do anything.

Begging the question

Can the referee simulate independent samples from p using the messages from the players?


39
271
11
77778
39
W1
W2
W3
W4
W5
2
18
11
3
21
R
17

Theorem

$\forall k, \ell < \log k$, there exists no SMP with ℓ bits of communication per player for distributed simulation over $[k]$ with any finite number of players. (Even allowing public-coin and interactive protocols.)

Theorem

$\forall k, \ell < \log k$, there exists no SMP with ℓ bits of communication per player for distributed simulation over $[k]$ with any finite number of players. (Even allowing public-coin and interactive protocols.)

Proof.

By contradiction, [...] pigeonhole principle [...]. □

Theorem

$\forall k, \ell \geq 1$, there exists a private-coin protocol with ℓ bits of communication per player for distributed simulation over $[k]$, with expected number of players $O(k/2^\ell \vee 1)$.

Algorithm 1 Distributed Simulation for $\ell = 1$: basic version

Require: $n = 2k$ players, each with an i.i.d. sample from unknown p

for $1 \leq i \leq n$ **do**

  players $(2i-1)$ and $2i$ send one bit: whether their sample is i.

$\mathcal{R}$ receives these $n = 2k$ bits $M_1, \dots, M_n$.

if exactly one of the bits $M_1, M_3, \dots, M_{2k-1}$ is equal to one, say the bit M_{2i-1}, and the corresponding bit M_{2i} is zero, **then** $\mathcal{R}$ outputs $\hat{X} = i$;

else $\mathcal{R}$ outputs $\perp$ (abort).

Algorithm 2 Distributed Simulation for $\ell = 1$: basic version

Require: $n = 2k$ players, each with an i.i.d. sample from unknown p

1: **for** $1 \leq i \leq n$ **do**

2: players $(2i-1)$ and $2i$ send one bit: whether their sample is i.

3: $\mathcal{R}$ receives these $n = 2k$ bits $M_1, \dots, M_n$.

4: **if** exactly one of the bits $M_1, M_3, \dots, M_{2k-1}$ is equal to one, say the bit M_{2i-1}, and the corresponding bit M_{2i} is zero, **then** $\mathcal{R}$ outputs $\hat{X} = i$;

5: **else** $\mathcal{R}$ outputs $\perp$ (abort).

Then $\forall i, \Pr[\hat{X} = i] = p_i \cdot (1 - p_i) \cdot \prod_{j \neq i}(1 - p_j) = p_i \cdot \prod_{j=1}^{k}(1 - p_j) \propto p_i$

Corollary (Informal)

For any inference task $\mathcal{P}$ over k-ary distributions with sample complexity s in the non-distributed model, there is a private-coin protocol for $\mathcal{P}$, with ℓ bits of communication per player, and $n = O(s \cdot k/2^{\ell})$ players.

Illustration ©Dami Lee

Corollary (Distribution Learning)

$\forall k, \ell \leq \log_2 k$, there is a *private-coin* protocol for learning k-ary distributions with ℓ bits per player, and $n = O(\frac{k^2}{2^\ell \varepsilon^2})$ players.

Corollary (Uniformity Testing)

$\forall k, \ell \leq \log_2 k$, there is a *private-coin* protocol for testing uniformity over $[k]$ with ℓ bits per player, and $n = O(\frac{k^{3/2}}{2^\ell \varepsilon^2})$ players.

Natural Question

Is this “simulate-and-infer” approach optimal?

Natural Question

Is this "simulate-and-infer" approach optimal?

Answer

Not if one allows public randomness!

“MINIMALLY CONTRACTING HASHING”

Theorem (Upper Bound)

$\forall k, \ell \leq \log_2 k$, there is a public-coin protocol for testing uniformity over $[k]$ with ℓ bits per player, and $n = O\left(\frac{k}{2^{\ell/2}\varepsilon^2}\right)$ players.

Theorem (ℓ_2 contraction)

Choose u.a.r. a balanced partition Π of $[k]$ in L parts, and let p_Π be the distribution induced by p on Π. Then

$$\Pr_{\Pi}[\ell_1(p_\Pi, u_L) \geq \Omega(\sqrt{L/k})\ell_1(p, u_k)] \geq \Omega(1)$$

Theorem (ℓ_2 contraction)

Choose u.a.r. a balanced partition Π of $[k]$ in L parts, and let p_Π be the distribution induced by p on Π. Then

$$\Pr_\Pi[\ell_1(p_\Pi, u_L) \geq \Omega(\sqrt{L/k})\ell_1(p, u_k)] \geq \Omega(1)$$

Proof.

Technical (and more general). Dealing with dependencies when computing second and fourth moments + Paley–Zygmund. □

Theorem (ℓ_2 contraction)

Choose u.a.r. a balanced partition Π of $[k]$ in L parts, and let p_Π be the distribution induced by p on Π. Then

$$\Pr_{\Pi}[\ell_1(p_\Pi, u_L) \geq \Omega(\sqrt{L/k})\ell_1(p, u_k)] \geq \Omega(1)$$

Proof.

Technical (and more general). Dealing with dependencies when computing second and fourth moments + Paley–Zygmund. □

(This is tight).

Apply with $L := 2^{\ell}$, choosing a common random Π using public coins.
Test p_Π with $\varepsilon' := \sqrt{L/k}\varepsilon$:

$$\frac{\sqrt{L}}{\varepsilon'^2} = \frac{\sqrt{k}}{2^{\ell/2}\varepsilon^2}\,.$$

Apply with $L := 2^{\ell}$, choosing a common random Π using public coins.
Test p_{Π} with $\varepsilon' := \sqrt{L/k}\varepsilon$:

$$\frac{\sqrt{L}}{\varepsilon'^2} = \frac{\sqrt{k}}{2^{\ell/2}\varepsilon^2}\,.$$

Repeat in parallel to amplify probability. $\square$

Algorithm 3 ℓ-bit public-coin protocol for uniformity testing.

Require: Parameter $\varepsilon \in (0,1)$, n players, each with an i.i.d. sample from unknown p

1: Set $L \leftarrow 2^{\ell}$

2: Players use independent public coins to sample a random partition $(S_1, \dots, S_L)$ of $[k]$ with equal-sized parts.

3: Upon observing the sample X_j, player j sends

$$M_j \leftarrow \sum_{b=1}^{L} b\mathbf{1}[X_j \in S_b]$$

(which part the sample fell in) ▷ $\log_2 L = \ell$ bits

4: $\mathcal{R}$ obtains n independent samples from $p' := (p(S_1), \dots, p(S_L))$ on $[L]$ and tests if p' is u_L or $(\varepsilon/\sqrt{L})$-far from uniform in ℓ_2 ▷ Uses a non-distributed ℓ_2 test.

5: $\mathcal{R}$ outputs what the ℓ_2 test outputs.

Algorithm 4 ℓ-bit public-coin protocol for uniformity testing.

Require: Parameter $\varepsilon \in (0,1)$, n players, each with an i.i.d. sample from unknown p

Set $L \leftarrow 2^\ell$

Players use independent public coins to sample a random partition $(S_1, \dots, S_L)$ of $[k]$ with equal-sized parts.

Upon observing the sample X_j, player j sends

$$M_j \leftarrow \sum_{b=1}^{L} b\mathbf{1}[X_j \in S_b]$$

(which part the sample fell in) ▷ $\log_2 L = \ell$ bits

$\mathcal{R}$ obtains n independent samples from $p' := (p(S_1), \dots, p(S_L))$ on $[L]$ and tests if p' is u_L or $(\varepsilon/\sqrt{L})$-far from uniform in ℓ_2 ▷ Uses a non-distributed ℓ_2 test.

$\mathcal{R}$ outputs what the ℓ_2 test outputs.

+ repeat in parallel.

- Simple.
- ℓ_2/χ^2 contraction theorem: very general.
- Randomness-hungry: $O(k\ell)$ random bits (Can improve to $O(\log k)$ using 4-wise independent only!)

With local communication constraints (upper bounds):

Task $\mathcal{P}$	n (private-coin)	n (public-coin)
Distribution learning	$\frac{k}{\varepsilon^2} \cdot \frac{k}{2^\ell}$	$\frac{k}{\varepsilon^2} \cdot \frac{k}{2^\ell}$
Uniformity testing	$\frac{\sqrt{k}}{\varepsilon^2} \cdot \frac{k}{2^\ell}$	$\frac{\sqrt{k}}{\varepsilon^2} \cdot \sqrt{\frac{k}{2^\ell}}$

CONCLUSION

in passing: that's optimal.

In different work ([ACT19], to appear in COLT'19), we provide a general lower bound framework.

- Framework for inference problems with communication constraints over discrete distributions: generalizes to other constraints [ACFT19]

- Framework for inference problems with communication constraints over discrete distributions: generalizes to other constraints [ACFT19]
- First work on distributed testing; optimal protocols for public-coin and private-coin uniformity testing in all settings considered

- Framework for inference problems with communication constraints over discrete distributions: generalizes to other constraints [ACFT19]
- First work on distributed testing; optimal protocols for public-coin and private-coin uniformity testing in all settings considered

CONCLUSION

- Framework for inference problems with communication constraints over discrete distributions: generalizes to other constraints [ACFT19]
- First work on distributed testing; optimal protocols for public-coin and private-coin uniformity testing in all settings considered
- Simple algorithms: should work well in practice?

CONCLUSION

- Framework for inference problems with communication constraints over discrete distributions: generalizes to other constraints [ACFT19]
- First work on distributed testing; optimal protocols for public-coin and private-coin uniformity testing in all settings considered
- Simple algorithms: should work well in practice?
- Many questions and directions to explore: several samples, functional estimation, more trade-offs…

Thank you

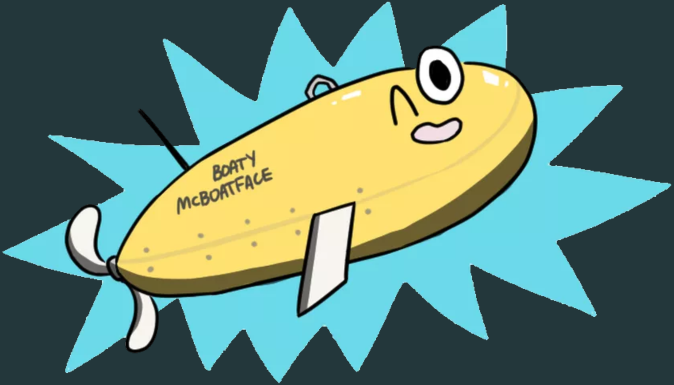


Illustration ©Dami Lee

Jayadev Acharya, Clément L. Canonne, Cody Freitag, and Himanshu Tyagi.
Test without Trust: Optimal Locally Private Distribution Testing.
AISTATS 2019, 2019.

Jayadev Acharya, Clément L. Canonne, and Himanshu Tyagi.
Lower bounds from chi-square contraction.
In Proceedings of COLT, 2019.

Ilias Diakonikolas, Elena Grigorescu, Jerry Li, Abhiram Natarajan, Krzysztof Onak, and Ludwig Schmidt.
Communication-efficient distributed learning of discrete distributions.
In Proceedings of NIPS, pages 6394–6404, 2017.

Yanjun Han, Pritam Mukherjee, Ayfer Özgür, and Tsachy Weissman.
Distributed statistical estimation of high-dimensional and nonparametric distributions with communication constraints, February 2018.
Talk given at ITA 2018.

Yanjun Han, Ayfer Özgür, and Tsachy Weissman.
Geometric lower bounds for distributed parameter estimation under communication constraints.
volume 75 of Proceedings of Machine Learning Research, pages 3163–3188. PMLR, 2018.

Or Sheffet.
Locally private hypothesis testing.
In Proceedings of the 35th International Conference on Machine Learning (ICML), volume 80, pages 4612–4621. PMLR, 10–15 Jul 2018.