# Lecture 2: Importance of Problem Structure

Lecturer: Yanjun Han

March 31, 2021

# Today's plan

Communication complexity of equality evaluation under:

- deterministic protocol
- randomized protocol with private randomness
- randomized protocol with public randomness
- randomized protocol with one-round communication

Idea:

- see how the lower bound is sensitive to slight changes in the problem
- no "skeleton key" for lower bounds

# Equality evaluation problem

Problem setup:

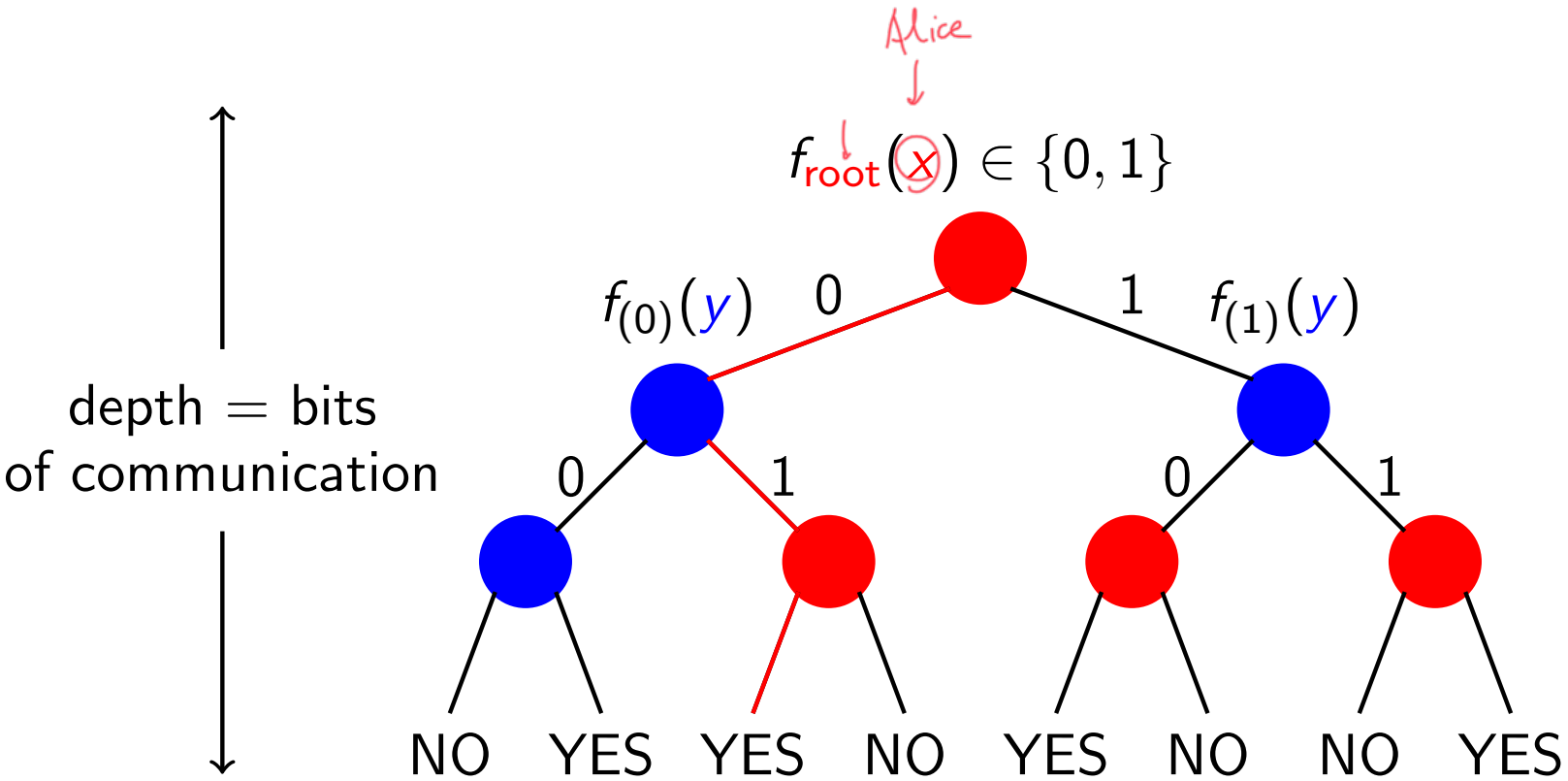
- Alice has a binary vector $x \in \{0,1\}^n$
- Bob has a binary vector $y \in \{0,1\}^n$
- they want to know whether $x = y$ or not
- however, Alice and Bob do not live together, and they need to communicate
- they are allowed to use any blackboard communication protocol

Simple scheme, $n+1$ bits

Target: under different performance metrics and resources, characterize the smallest number of bits required to communicate, a.k.a. the communication complexity.

# Blackboard communication protocol

Red - Alice, Blue - Bob




message = 010, output = YES

# Setting I: deterministic communication complexity

Target: for each input pair $(x, y)$, the protocol always outputs the correct answer; i.e.

output of protocol

$$\forall x, y \in \{0,1\}^n : \mathbb{P}(P(x, y) = 1(x = y)) = 1.$$

**Theorem**

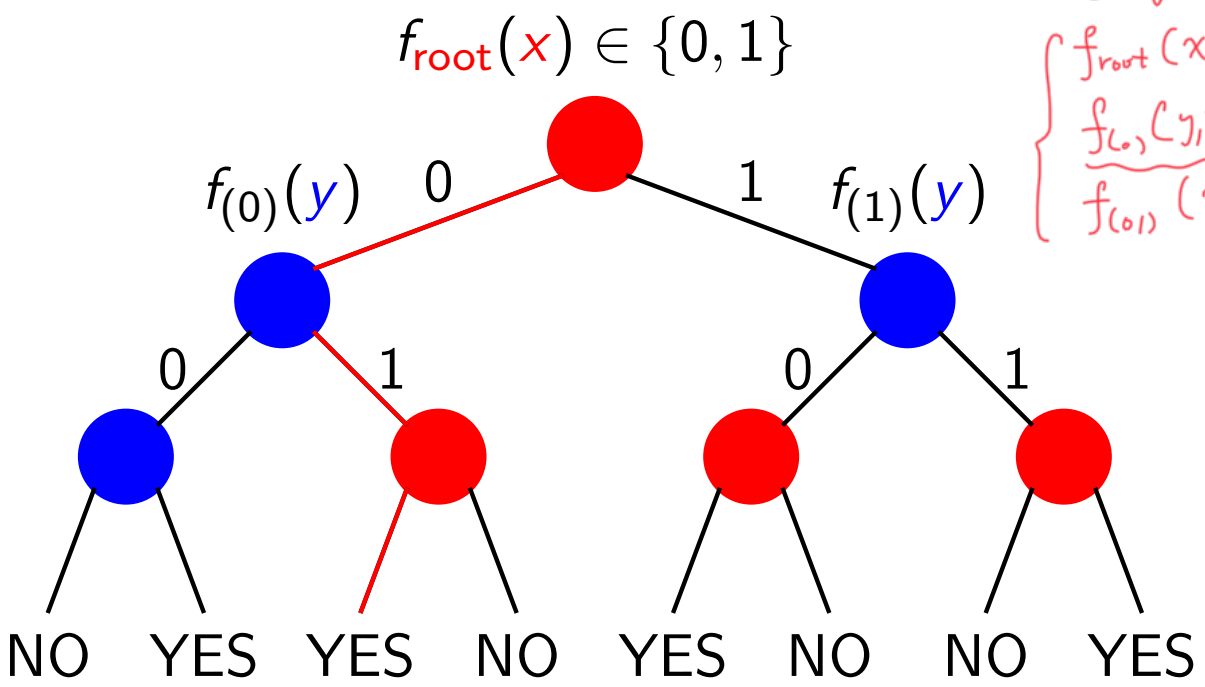
The deterministic communication complexity of equality evaluation is $n + 1$ bits.

An easy upper bound: Alice $\xrightarrow{x}$ Bob, Bob $\xrightarrow{1(x=y)}$ Alice

# Copy-paste property

**Lemma**

For a deterministic blackboard protocol, if inputs $(x_1, y_1)$ and $(x_2, y_2)$ arrive at the same leaf node, then $(x_2, y_1)$ and $(x_1, y_2)$ also go to this node.




$(x_1, y_1)$ $(x_2, y_2)$

$f_{root}(x_1) = 0 = f_{root}(x_2)$

$f_{(0)}(y_1) = 1 = f_{(0)}(y_2)$

$f_{(01)}(x_1) = 0 = f_{(01)}(x_2)$

# Rectangle and rank

Corollary: for each leaf node $v$, the set $P_v$ of all input pairs going to $v$ is a rectangle, i.e. $P_v = X_v \times Y_v$.

$(x,y)$ goes to $v$ iff $x \in X_v$, $y \in Y_v$

**Theorem (Log-rank inequality)**

Let $P \in \{0,1\}^{X \times Y}$ be a matrix defined as $P(x,y) = f(x,y)$ for all $x \in X, y \in Y$. Then the deterministic communication complexity of computing $f$ is at least $\log_2 \operatorname{rank}(P)$. → over $\mathbb{R}$

For equality: $f(x,y) = 1(x=y)$ $P \in \mathbb{R}^{2^n \times 2^n}$ $P = I_{2^n}$

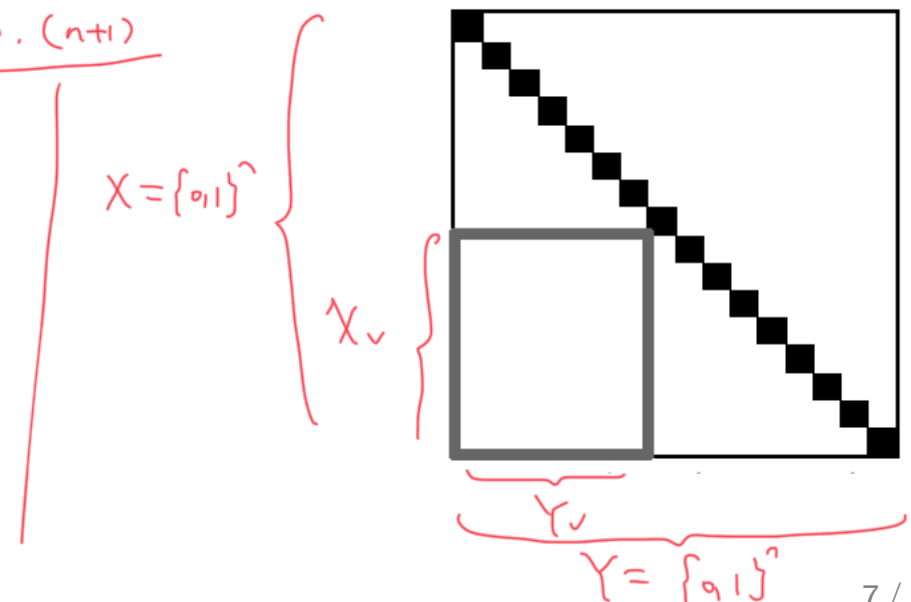
$\operatorname{Rank}(P) = 2^n$ Theorem $\Rightarrow$ DCC $\geq n \cdot (n+1)$

Proof: $P = \sum_v P_v$, $v \in$ {leaf nodes corresponding "YES"}

$$P_v(x,y) = \begin{cases} 1, & (x,y) \text{ goes to } v \\ 0, & \text{o.w.} \end{cases}$$

$\operatorname{rank}(P) \leq \sum_v \operatorname{rank}(P_v) =$ #{leaf nodes corresponding "YES"}

$\Rightarrow$ depth of tree $\geq \log_2 \operatorname{rank}(P)$



# Setting II: randomized complexity with private coin

Additional resources: Alice and Bob have some private randomness sources $R_A$ and $R_B$ $\quad R_A \perp\!\!\!\perp R_B$

Target: for each input pair $(x, y)$, the protocol outputs the correct answer with high probability; i.e.

$$\forall x, y \in \{0,1\}^n : \mathbb{P}_{R_A, R_B}(P(x,y) = 1(x = y)) \geq 0.99.$$

**Theorem**

The randomized communication complexity of equality evaluation with private coins is $\Theta(\log n)$ bits.

$\hookrightarrow a_n = \Theta(b_n) \iff \exists\, 0 < c_1 < c_2 < \infty$ s.t. $c_1 b_n \leq a_n \leq c_2 b_n$

$\implies 0 < \liminf_n \frac{a_n}{b_n} \leq \limsup_n \frac{a_n}{b_n} < \infty.$

Which part of the previous lower bound breaks down?

# Upper bound

A cute scheme by Rabin and Yao:

$\log_2(2n^2) = O(\log n)$

$(p, t, P(x,t))$

Alice

$x = (x_0, \cdots, x_{n-1})$

finds a prime number

$p \in [n^2, 2n^2]$

generate $t \in \{0, 1, \cdots, p-1\}$ u.a.r.

evaluate polynomial

$P(x,t) = x_0 + x_1 t + x_2 t^2 + \cdots + x_{n-1} t^{n-1} \pmod p$

Bob

$y$

checks whether

$P(y,t) = P(x,t)$

---

Correctness:

$x = y$ scheme always outputs "YES"

$x \neq y$ $\mathbb{P}(P(y,t) = P(x,t))$ "small" $\leq \frac{n-1}{p} < \frac{1}{n}$.

$\downarrow$

$(x_0 - y_0) + (x_1 - y_1)t + \cdots + (x_{n-1} - y_{n-1})t^{n-1} = 0 \pmod p$

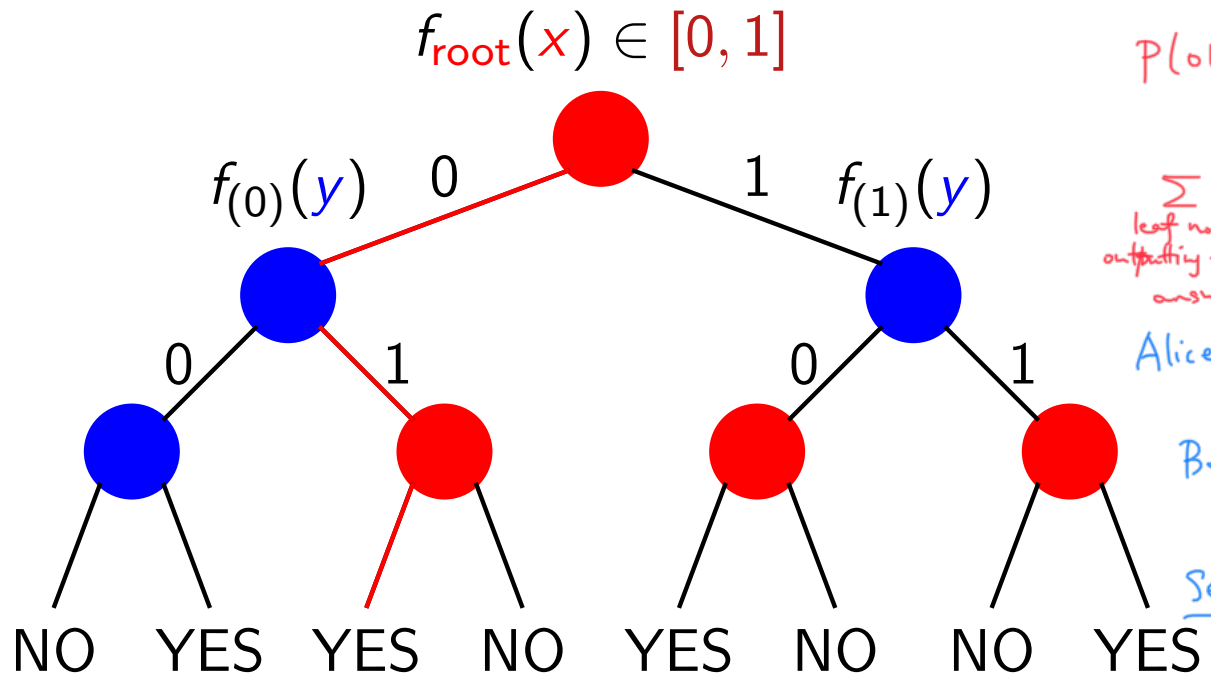
non-zero poly. of $t$ w. deg. $\leq n-1$

# Lower bound

## Theorem

Let $D(f)$ and $R(f)$ be the deterministic and randomized private-coin communication complexity of computing $f$, respectively. Then

$$D(f) \leq 2^{R(f)} \left( R(f) - \log_2 \left( \frac{1}{2} - \varepsilon \right) \right).$$




$P(010) = (1 - f_{root}(x)) \cdot f_{(0)}(y) \cdot (1 - f_{(01)}(x))$

$= P_A(010) \cdot P_B(010)$

$\sum_{\text{leaf node } v \text{ outputting the correct answer}} P(v) \geq 1 - \varepsilon.$

Alice sends $\{P_A(v) : \text{all leaf node } v\}$, card. $2^{R(f)}$

Bob computes $P(v)$ for every $v$

majority vote

Sending real number: quantize into accuracy level $2^{-R(f)} \cdot (\frac{1}{2} - \varepsilon)$

# Setting III: randomized complexity with public coin

Additional resources: Alice and Bob have public randomness source $R$

Target: for each input pair $(x, y)$, the protocol outputs the correct answer with high probability; i.e.

$$\forall x, y \in \{0,1\}^n : \mathbb{P}_R(P(x,y) = 1(x = y)) \geq 0.99.$$

**Theorem**

The randomized communication complexity of equality evaluation with public coins is $\Theta(1)$ bits.

Generate random $v_1, v_2, \dots, v_{100} \in \mathbb{F}_2^n$

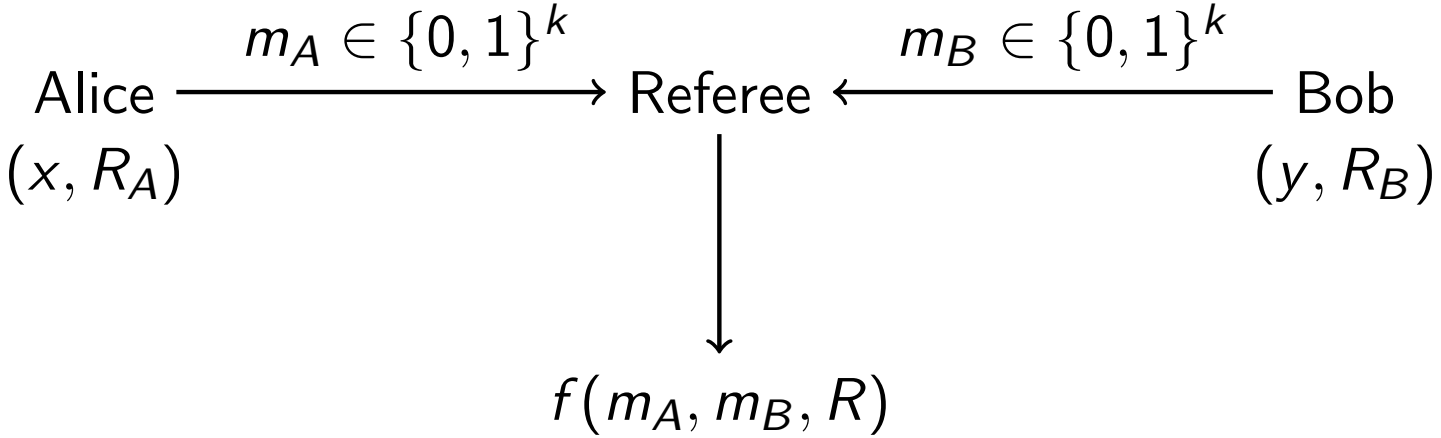
Alice sends $\langle v_1, x\rangle, \langle v_2, x\rangle, \dots, \langle v_{100}, x\rangle$

Bob checks whether $\langle v_i, x\rangle = \langle v_i, y\rangle$ for all $i \in [100]$

$\mathbb{P}(\langle v_i, x\rangle = \langle v_i, y\rangle) = \frac{1}{2}$

Which part of the previous lower bound breaks down?

# Setting IV: one-round complexity with private coin




Target: it holds that

$$\forall x, y \in \{0,1\}^n : \mathbb{P}_{R_A,R_B,R}(f(m_A, m_B, R) = 1(x = y)) \geq 0.99.$$

**Theorem**

The one-round communication complexity of equality evaluation with private coins is $\Theta(\sqrt{n})$ bits.

# Lower bound: characterizing all strategies

Notation:

- $P_A(m_A \mid x)$: prob. of Alice sending $m_A$ when holding $x$
- $P_B(m_B \mid y)$: prob. of Bob sending $m_B$ when holding $y$
- $P_R(1 \mid m_A, m_B)$: prob. of Referee outputting YES under $m_A, m_B$

for each $(x, y) \in \{0,1\}^n \times \{0,1\}^n$,

$$\left| \sum_{m_A} \sum_{m_B} P_R(1 \mid m_A, m_B) \cdot P_A(m_A \mid x) \cdot P_B(m_B \mid y) - \mathbb{1}(x = y) \right| \leq 0.01.$$

$\Downarrow$

$P_A \in \mathbb{R}^{2^n \times 2^k}$ $P_R \in \mathbb{R}^{2^k \times 2^k}$ $P_B \in \mathbb{R}^{2^n \times 2^k}$

$$\| P_A P_R P_B^T - I_{2^n} \|_\infty \leq 0.01$$

↑ not matrix norm

**Theorem (Newman and Szegedy'96)**

If there exist stochastic matrices $A, B \in \mathbb{R}^{N \times m}$ and a square matrix $R \in \mathbb{R}^{m \times m}$ such that $\|ARB^\top - I_N\|_\infty \leq 0.1$. Then $m = 2^{\Omega(\sqrt{\log N})}$.

$N = 2^n$, $m = 2^k$ $\quad m = 2^{\Omega(\sqrt{\log N})} \Longrightarrow k = \Omega(\sqrt{n})$

# Upper bound: simulation

The idea of simulation (Newman'91):

- Let $R$ be a public-coin protocol with error probability $\leq \varepsilon$ on all inputs
- Draw $m$ fixed iid copies $R_1, \cdots, R_m$, then if $m = Cn$ with $C$ large enough, we have

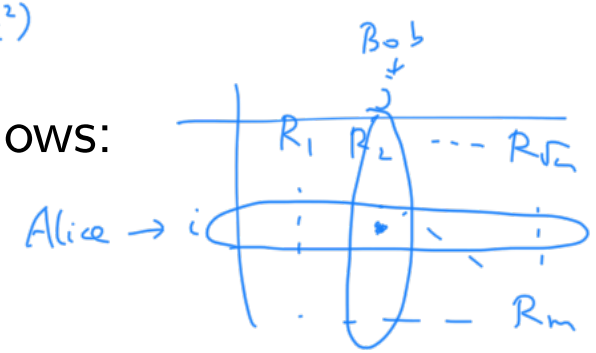
$$\mathbb{P}\left(\exists x, y \in \{0,1\}^n : \frac{1}{m}\sum_{i=1}^{m} 1(R_i(x,y) \neq 1(x=y)) \geq 2\varepsilon\right) < 1.$$

$\leq 2^{2n} \cdot \max_{x,y} \mathbb{P}\left(\frac{1}{m}\sum_{i=1}^{m} 1(R_i(x,y) \neq 1(x=y)) \geq 2\varepsilon\right)$ $\leq 2^{2n} \cdot e^{-2m\varepsilon^2}$ $m = \frac{Cn}{\varepsilon^2}$

$\mathbb{E} \leq \varepsilon$, $|\cdot| \leq 1$ $= \exp.\ \text{small}.$

Hoeffding $\Rightarrow$ $\exp(-2m\varepsilon^2)$

- Interleave the above $R_1, \cdots, R_m$ as follows:




complexity $= O(\sqrt{m}) = O(\sqrt{n})$

# Summary

Communication complexity of equality evaluation under:

- deterministic protocol ($\Theta(n)$ bits)
- randomized protocol with private randomness ($\Theta(\log n)$ bits)
- randomized protocol with public randomness ($\Theta(1)$ bits)
- randomized protocol with one-round communication ($\Theta(\sqrt{n})$ bits)

# Reading materials


- Ilan Newman, and Mario Szegedy. "Public vs. private coin flips in one round communication games." In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pp. 561–570, 1996.
- Alexander Razborov. "On the distributed complexity of disjointness." *Theoretical Computer Science*, 106: 385-390, 1992.
- Eric Blais, Clément L. Canonne, and Tom Gur. "Distribution testing lower bounds via reductions from communication complexity." *ACM Transactions on Computation Theory* 11, no. 2 (2019): 1-37.


Next lecture: $f$-divergence, joint range, statistical decision theory