Today: Average Case Complexity: Definitions.
- Distributional Problem?
- Feasible Problems?
- Intractable Ones?
- Reductions?

Based entirely on:
Oded Goldreich: Conceptual Intro To Computational Complexity Sets 10.2
Question:
- Is TSP hard on average or easy?

Answer:
- Depends who you ask!
- If we pick points uniformly from an \( n \times n \) square ... then seems easy.
- if you pick entire grid & perturb each point a bit, then seems hard.

Conclusion:
Complexity is a function of
(i) Problem &
(ii) Distribution.
Distributional Problems

Specified by a pair \( (\Pi, D) \)

\[ \Pi \subseteq \{0,1\}^* \times \{0,1\}^* : \text{usual relational problem} \]

\[ D = \{ D_n \} : D_n : \{0,1\}^n \rightarrow [0,1] \text{ is a distribution on } \{0,1\}^n. \]

---

Goal: Given \( \mathcal{D} \leftarrow_D \{0,1\}^n \)

find \( y \) s.t. \( (x, y) \in \Pi \).

---

Complexity Measure?

Expected running time? Not so interesting
Examples:

1. Suppose $A$ solves $\Pi$ on $D$ as follows:
   - w.p. $2^{-\frac{\epsilon}{n}}$, $A$ takes time $2^n$.
   - w.p. $1 - 2^{-\frac{\epsilon}{n}}$, $A$ takes time $n^2$.

   Is this "polynomial"? Exponential?

2. Suppose $B$ solves $\Pi'$ on $D'$ as follows:
   - w.p. $\sim \frac{1}{c^2}$, $B$ takes time $n^c$.

   Is this "polynomial"?
Our Preference

Avg-Time = "Time as viewed by polytime observer",

& not

"What could be sensed after unreasonable sampling"

Back to Examples:

(i) In any poly # samples, very unlikely to see exponential behavior.

⇒ Avg-Time = \( n^2 \).

(ii) For every \( c \), prob. of seeing run time \( \geq n^c \), io \( \geq \frac{1}{c^2} \).

⇒ Avg. Time = super-poly.
**Formal Definition:**

$\text{Avg-Timu of } A \text{ on } (\Pi, D) \text{ is } \leq T(n)$

if \( \forall n, c \)

\[ \Pr \left[ \begin{array}{l}
\text{A}(x) \text{ incorrect} \\
\text{x} \leftarrow_{D_n} \{0, 1^n\} \text{ or A(x) runs in time} \\
\geq T(n) \end{array} \right] \leq \frac{1}{n^c}. \]

---

**Note:** Allowing $A$ to be incorrect makes definitions equivalent.

$\text{Avg-BPP} = \left\{ (\Pi, D) \mid \exists A, c \text{ solving (\Pi, D)} \text{ in Avg-Timu } n^c \right\}$. 
**Intractable Problems?**

**Attempt 1:**

\[ \text{DNP}_1 = \{ (\pi, d) \mid \pi \in \text{NP}, \text{ (i.e., "}(x,y) \in \pi?\text{" decidable in P)} \} \]

**Problem**

- Notion of distribution too strong, for "empirical" concerns.
- Can easily prove:

\[ \text{NP} \not\equiv \text{BPP} \Rightarrow \text{DNP}_1 \not\equiv \text{Avg. BPP} \]

Worst-case hardness \(\Rightarrow\) average case hardness.
- $D_{A,n}^{\text{Adv}}$ uniform on $\frac{1}{2}x \mid A(x)$

- $D_{n}^{\text{Adv}} \leq \frac{1}{i} D_{A_i,n}^{\text{Adv}}$

$A_1, A_2, \ldots A_i, \ldots$ Enumeration of BPP w/c.

- Diagonalization by Distribution!

- Problem: Distributions worse than adversary, which avg case wants to understand "naturally-occurring instances".

- Shouldn't allow arbitrary distributions $D$. 
Sampleable Distributions

Model of Universe

- What kind of distribution do we see?
  - "Sampleable Distribution"
Definition: \( D \) is sampleable if

\[ \exists \text{ poly-time (deterministic) algorithm } g \]

\[ \forall n, x \in \{0,1\}^n \]

\[ \Pr_y \left[ g(y) = x \right] = D(x) \]

\( y \leftarrow \text{ uniform on } \{0,1\}^n \)

---

Interesting Intractable Problems

\[ \mathcal{DNP} = \left\{ (\Pi, D) \mid \Pi \in \mathcal{NP}, \exists D \text{ sampleable} \right\} \]
Basic Questions

- Is $\text{DNP} \leq \text{Avg. BPP}$? No.
- If $\text{NP} \cap \text{BPP}$ then, Yes, but is $\text{DNP} \cup \text{Avg. BPP}$? Can't prove.
- Find some "worst-case assumption" that implies $\text{DNP} \\cup \text{Avg. BPP}$.
- What are some $\text{DNP}$-complete problems?
- What is completeness? reductions?
Replacements

- **Deterministic**: Simple ... should help solve original problem.

- **Probabilistic**: Already get complex ... needn't always be correct.

- **Distributional**: Trickier ... can be incorrect; can produce unlikely instances.
More formally

Most restrictive notion:

- (Deterministic Reduction): \((R, T)\) reduce

\[(\Pi_1, D_1) \rightarrow (\Pi_2, D_2)\]

if

(i) \(R, T\) are polytime.

(ii) \((R(x), y) \in \Pi_2\)

\[\Rightarrow (x, T(y)) \in \Pi_1\]

(iii) \(R(x)\) distributed as \(D_2\)

if \(x\) distributed as \(D_1\).
But don't need to adhere to distribution so stringently.

- Alg for $\Pi_2$ doesn't "know" $D_2$.

- Domination of Distributions

- $D_2 \alpha$-dominates $D_1$ if

  \[ \forall x \quad D_1(x) \leq \alpha(D_1(x), D_2(x)). \]

  (Pictorially)
(Weaker def. reduction)

\[ (\Pi_1, D_1) \rightarrow (\Pi_2, D_2) \]

\[ a \xrightarrow{R} R(x) \]

\[ T(y) \xleftarrow{T} y \]

\[ a \leftarrow D_1 \rightarrow R(x) \text{ drawn from } D_2' \]

\[ \text{s.t. } D_2 \text{ poly. dominates } D_2' \]

Claim: Such a reduction \( (\Pi_2, D_2) \in \text{Avg.\ BPP} \)

\[ \Rightarrow (\Pi_1, D_1) \in \text{Avg.\ BPP} \]
Can also consider randomized...

- Definition simpler.
- Will see example next lecture.