**2.5 Multi-Armed Bandits in Healthcare**

Prof Cem Tekin (Bilkent University)
http://cyborg.bilkent.edu.tr/

**MAB introduction**

- MAB problem
  - Gambling in a K slot machines over rounds
  - In each round:
    - Play an arm
    - Observe and collect its random reward
  - Goal:
    - Maximize expected cumulative reward
- Adaptive clinical trials:
  - Arms – treatments
  - Sequentially allocating treatments to patients
  - Goal:
    - Correctly identify best treatment (exploration)
    - Effectively treat as many patients in clinical trials as possible (exploitation)
  - Gittins & jones
    - Bayesian approach
    - https://www.jstor.org/stable/2335176
- Bolus insulin recommendation
  - Arms are insulin doses
  - Sequentially administer bolus insulin to patient for blood glucose regulation
  - Goal:
    - Keep blood glucose as close as possible to a target level as long as possible:
      - Exploration: learn effectiveness of different doses
      - Exploitation: learnt optimal dose
      - Safety: even a single bad recommendation can have severe consequences
- Basic MAB model
  - MAB environment:
    - Arm (treatment/dose) set K = {1,....K)
  - Environemnt class
    - E.g. all K-armed bandits with Bernouli rewards
      - For binary outcomes
    - E.g. K-armed bandits with rewards in [0,1]
- How to choose $A_t$?
  - History = $H_t = \{A_1, R_1, .... A_{t-1}, R_{t-1}\}$
- Regret of a policy
  - Goal maximize: highest cumulative reward over T rounds
  - Maximising rewards = minimising regrets
- What is good policy?

- Regret lower bound
    - Consistent policy:
        - Not all policy is consistent
    - Asymptotic lower bound (Lai and Robbins 1985)
        - https://www.sciencedirect.com/science/article/pii/0196885885900028

**Principle of optimism**

- Lai and Robbins 1985:
    - Expected arm reward not known
    - Compute an (over)estimate such than
    - Select $A_{t,}$ = arg $\max_a \hat{u}_a(H_t)$#
- An instantiation of optimism: UCB policy
    - https://homes.di.unimi.it/~cesabian/Pubblicazioni/ml-02.pdf

**UCB in action**

- More playing, reduced confidence intervals

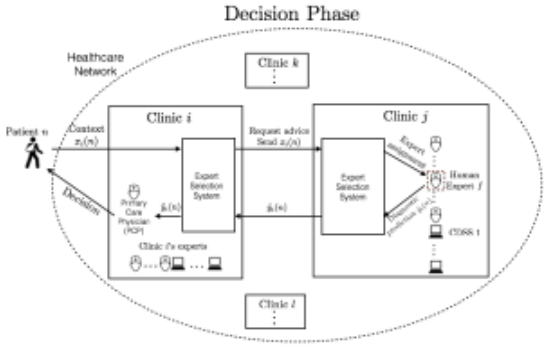
**Summary of Part 1**

- MAB model introduction
- Any consistent policy incurs at least log *T* regret
- UCB achieves log*T* regret

**Part2: Contextual MAB & Healthcare applications**

**Context MAB**

- Same arm but reward distributed according to context
- Contextual MAB: Personalized treatment
    - A patient visits with context including symptoms, labs, tests, etc.
    - Treatment is administered ($A_t$)
    - Patient response to treatment ($R_t$)
    - What is the best treatment for current patients
- Matching patients with experts




    - https://www.vanderschaar-lab.com/papers/Tekin_TETC2015.pdf
- Contextual MAB
    - https://proceedings.mlr.press/v15/chu11a.html

    - Contextual zooming:
        - https://arxiv.org/abs/0907.3986
    - Context gaussian process bandit:
        - https://papers.nips.cc/paper/2011/file/f3f1b7fc5a8779a9e618e1f23a7b7860-Paper.pdf
- Utilizing sparsity:
    - https://tor-lattimore.com/downloads/book/book.pdf
    - https://ieeexplore.ieee.org/document/7039192

**Safe leveling**
https://arxiv.org/abs/2111.13415

**TACO**

https://github.com/jxx123/simglucose