DMU

Probabilistic

$$P(A) = \sum_b P(A|b)P(b)$$

$$P(A,B)$$

$$P(A|B) = \frac{P(A,B)}{P(B)}$$

Bayes Rule

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Independence

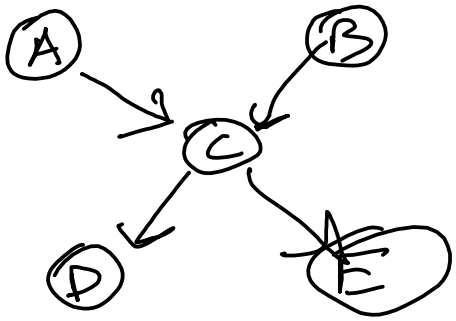
$$A \perp B \mid C$$

$$P(A,B|C) = P(A|C)P(B|C)$$

Stochastic Processes

$$\{X_t\}_{t=0}^{\infty}$$

Bayes Net



d-seperation $\iff$ Conditional independence

Sampling : Topological

Inferrence : NP-hard

Direct sampling

Likelihood-Weighted

Gibbs

Learning

Parameter

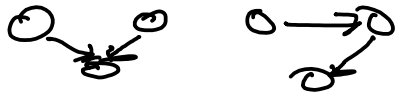
Max Likelihood

Bayesian

Non-parametric

Structure

Bayesian Score



K2

Local Optimization

Sequential Decision Problems

Outcome	Model	State
A/E	Static E	Dynamic E

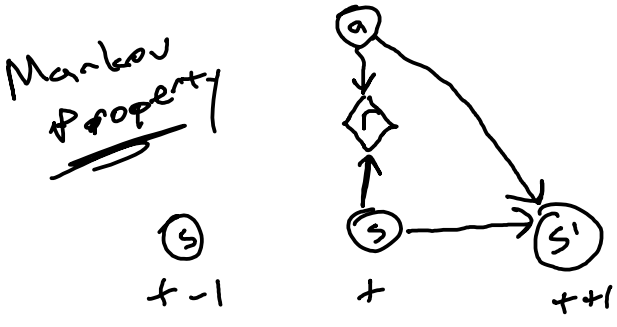
A

MDP

Generative/Explicit

(S, A, T, R, γ)

Markov property



maximize $\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t)$

deterministic policy on state

Offline

Policy Iteration

Value Iteration ← Contraction Mapping

$$V^*(s) = \max_a \{ R(s,a) + \gamma E[V^*(s') | s,a] \}$$

Approximate DP

Direct Policy Search

$$\pi(s) = f_\theta(s)$$

↑ choose

- Genetic
- Simulated Annealing
- Cross Entropy

Online MDP Methods

Forward Search $O(|A| \times |S|)^d$

Branch + Bound

Sparse Sampling ← continuous S

MCTS

Search Expansion Rollout Backup

↑ UCB

$$\text{argmax}_a \; Q(s,a) + c \sqrt{\frac{\log N(s)}{N(s,a)}}$$

→ UCT

Model Uncertainty (RL)

Model Based / Model Free

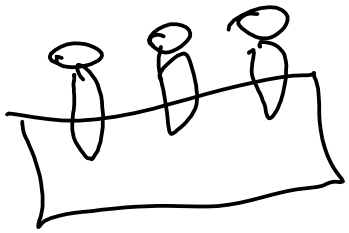
On Policy / Off Policy

Deep / Trad/Tab

Challenges.

- Exploration/Exploit
- Credit Assignment
- Generalization

Bandit



ε-greedy
softmax $e^{\beta Q}$

$O(\log)$ regret
- UCB
- Thompson
- Bayes Optimal
- Info Gathering

$\arg\max_a \frac{\Delta(a)^2}{g(a)}$

Tabular

Max Likelihood Model

On-Policy SARSA

Off Policy Q-learning

Deep RL

DQN — Replay Buffer, Frozen Target — RAINBOW

Policy Gradient — Baselines — TRPO, PPO

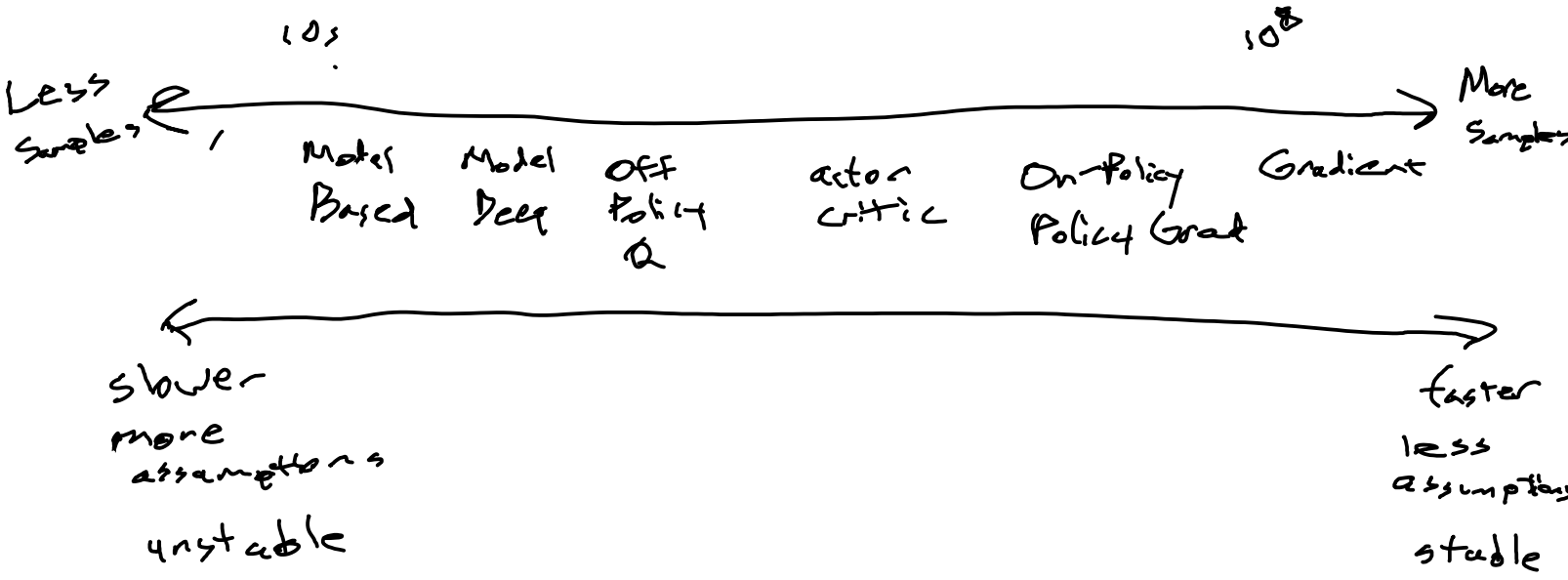
Actor-Critic ≃ Policy Gradient + Value Learning — A3C, GAE, SAC

Exploration : RND

Continuous A : DDPG

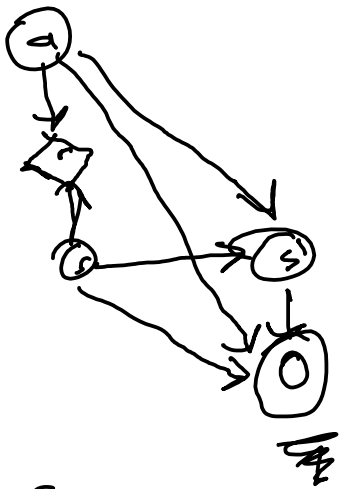
Transfer

Meta :



Inverse RL : Given Data Learn R

State Uncertainty (POMDPs)



$(S, A, O, T, Z, R, \gamma)$

$\pi : \mathcal{H} \rightarrow A$

Belief Update

Exact Bayesian

Particle Filtering

Offline POMDP

- α-vector VI
  - prune dominated
- PBVI
- SARSOP

Online POMDP

- AEMS
- MCTS/POMCP
- DESPOT
  - POMCPOW/DESPOT-α

Model Approximation

- Certainty Equivalence $\pi(E[s])$
- QMDP $\operatorname{argmax}(E[Q_{MDP}(s,a)])$

Games

- Equilibria
  - Nash
  - Dominant Strategy
- Differential Game

Alt Optimization Obj

- Multi Obj
  - Weighted Combination
  - Constrained
- Stochastic Policies
- Coherent Risk Measures

AlphaZero
AlphaStar
Deepstack