TUESDAY FEB 2, 2021

4 Inference

Bayes - Posterior - MPE - MAP
- L enum
- L sampling

5 Exact Inference

- Bayes $\boxed{\text{Filter}}$
- HMM → FACTOR GRAPHS

- MAX PRODUCT fn MPE
- SUM PRODUCT fn - posterior
  - MAP

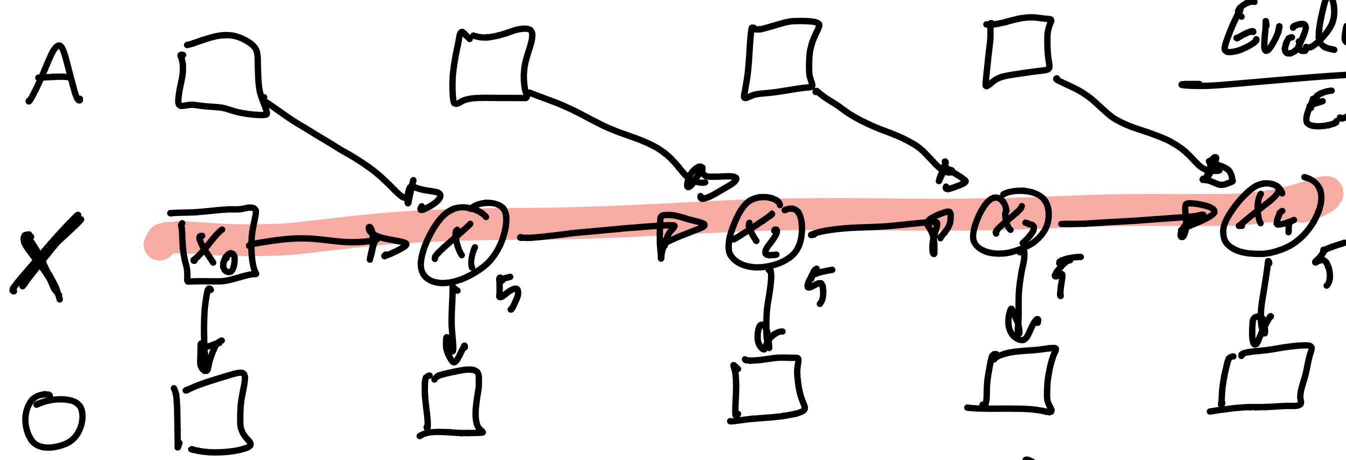
---

4 Inference

Sense → think → act

Evaluation

EASY



- Posterior $P(X \mid A, O, X_0)$
  - → 625 numbers
  - $5^{100} = 8\ 10^{69}$

→ $|X| = 5^4$
$= |X_t|^T$
$= 625$

- Max Probable Explanation

$\boxed{\text{Importance Sampling.}}$

$\pi(X) = P(X|A,O)$ TARGET POSTERIOR $\quad q(X) = P(X|A)$

$\forall i$

$$w^{(i)} = \frac{\boxed{\pi(X^{(i)})}}{\boxed{q(X^{(i)})}}$$ importance weight

proposal. $P(X|A)$ prior given Actions

1M

$X^{(i)}, A^{(i)}, O^{(i)}$

oversample → < 1.0

undersample > 1.0

$$\boxed{w^{(i)} = \frac{P(X|AO)}{P(X|A)}} \propto \frac{P(XAO)/\cancel{P(AO)}}{P(X|A)}$$

Fixed

$$\propto \frac{P(O|X)\cancel{P(X|A)}\cancel{P(A)}}{\cancel{P(X|A)}}$$

Fixed

Do N times:

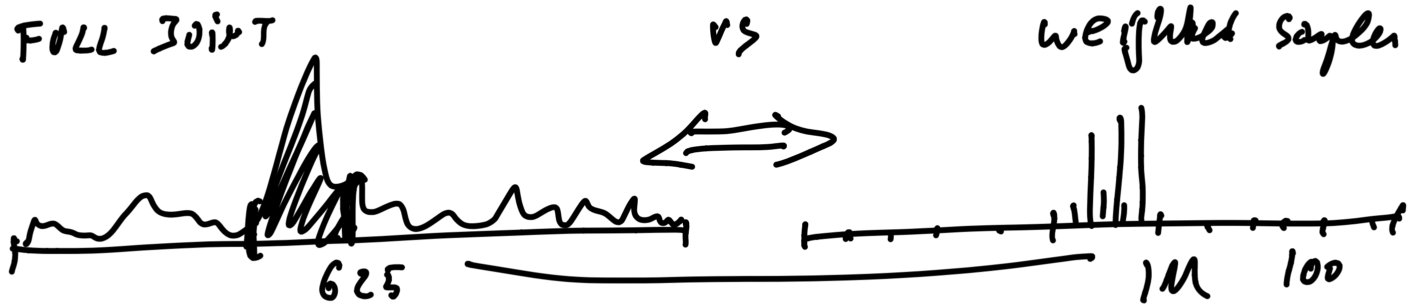
1) Sample $X^{(i)}$ From Markov Ch. ⟸

2) "Score" with $w^{(i)} = P(O|X)$

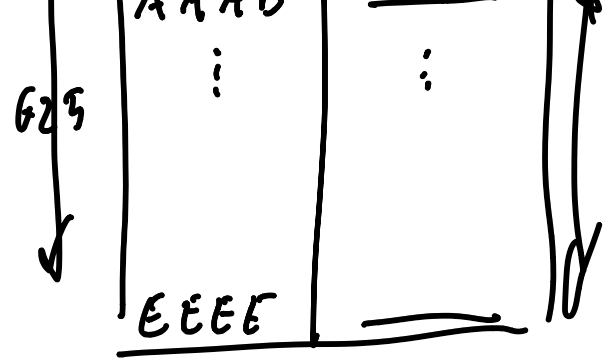
$$= P(O|X) = \prod P(O_t|X_t)$$

i.e. weight equals probability of measurements given states = very intuitive!

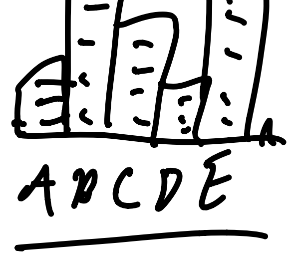
Downside: many weights $w^{(i)}$ will be close to 0: HIGH VARIANCE



A A A A
A A A B



$F(X) = X_5$



$F(X) = \#E$

$= \frac{1}{100}\sum_{i=1}^{100} \#E(X^{(i)})$

= MONTE CARLO ESTIMATE of any $F(X)$

HIGH VARIANCE if $|X| \gg$

BAYES LAW.

$P(X,Y) = P(X|Y)P(Y)$

$= P(Y|X)P(X)$

$P(X)$	X Y	FULL JOINT	$P(Y)$
.4	0 0	.2	.5
	0 1	.2	
.6	1 0	.2	.5
	1 1	.4	

$P(X|Y=0)$

$\underbrace{P(X|Y)}_{\text{post.}} = \frac{P(Y|X)\overbrace{P(X)}^{\text{prior}}}{P(Y)}$ BL 1

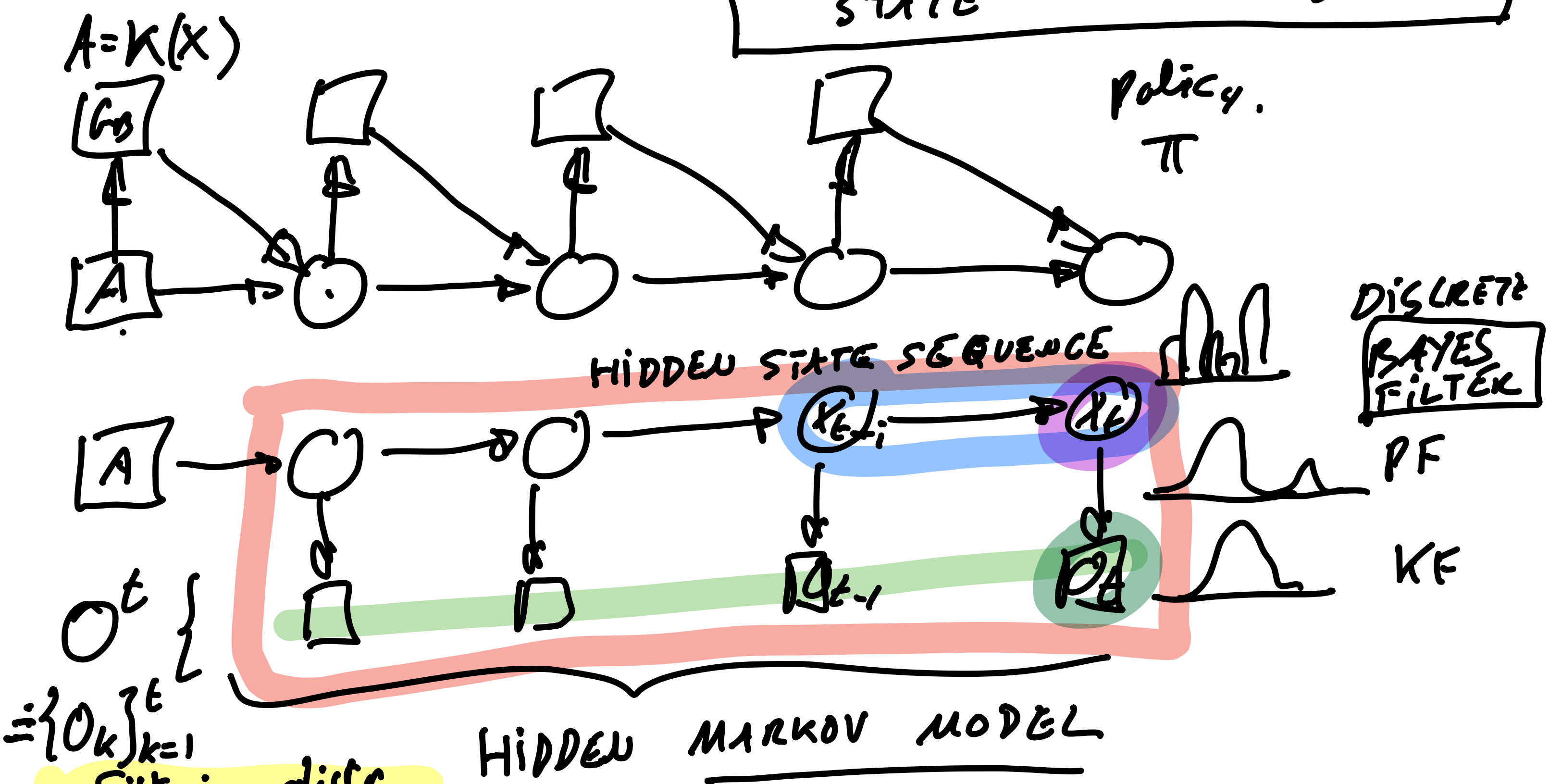
$P(X|Y) \propto \underbrace{P(Y|X)}_{\text{likelihood of X given Y}} \underbrace{P(X)}_{\text{prior}}$ BL 2

$\boxed{\underbrace{P(X|Y)}_{\text{posterior}} \propto \underbrace{L(X;Y)}_{\text{lik}} \underbrace{P(X)}_{\text{prior}} \quad \text{BL 3}}$

$L(X;Y) \propto P(Y|X)$

5 Exact Inference

inference over
- Sequence — DISCRETE / CONTINUOUS TRAJECTORY — SMOOTHING
- CURRENT STATE — FILTER — CURRENT STATE



Filtering distr.

$$P(X_t \mid O^t) = P(X_t \mid O_t, O^{t-1})$$

$$\propto L(X_t; O_t)\underbrace{P(X_t \mid O^{t-1})}_{\text{predictive distr.}}$$

$$\propto \underbrace{L(X_t; O_t)}_{\propto P(O_t \mid X_t)} \sum_{X_{t-1}} P(X_t \mid X_{t-1}) \underbrace{P(X_{t-1} \mid O^{t-1})}_{\text{filtering distr. at } (t-1)}$$

1) calc. predictive : → predict
2) multiply w likelihood → update