From previous class :

- Let's do this mathematically now. We want $\underline{P(s_k \mid m_{1:n})}$ ← offline
- Let's start with a basic result for $\underline{P(s_{1:n} \mid m_{1:n})}$ → trajectory given all measurements

$$P(s_{1:n} \mid m_{1:n}) = \frac{P(s_{1:n}, m_{1:n})}{P(m_{1:n})} \propto \underline{P(s_{1:n}, m_{1:n})}$$

$$\underbrace{P(s_{1:n}, m_{1:n})}_{\mathbb{R}^{2n}} \overset{\text{chain rule}}{=} P(m_n \mid m_{1:n-1}, s_{1:n})\, P(m_{n-1} \mid m_{1:n-2}, s_{1:n}) \cdots$$

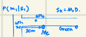
$$\cdots P(m_1 \mid s_{1:n})\, P(s_n \mid s_{1:n-1}) \cdots P(s_2 \mid s_1)\, P(s_1)$$

$P(ABC) = P(A|BC)\,P(B|C)\,P(C)$

Markovian

$$= P(m_n \mid s_n)\, P(m_{n-1} \mid s_{n-1}) \cdots P(m_1 \mid s_1)\, P(s_n \mid s_{n-1}) \cdots P(s_2 \mid s_1)\, P(s_1)$$

$$\boxed{\underbrace{P(s_{1:n}, m_{1:n})}_{\text{which trajectory \& measurements are in agreement.}} = P(m_1 \mid s_1)\, P(s_1) \prod_{i=2}^{n} \underbrace{P(m_i \mid s_i)}_{\text{sensor error}}\, \underbrace{P(s_i \mid s_{i-1})}_{\text{Transition prob. matrix}}}$$



$P(m_1 = GPS_1 \mid S_k = McD)$ = small

$P(m_1 = GPS_2 \mid S_k = McD)$ = large

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

- Now we want $\underline{P(s_k \mid m_{1:n})}$ → what is my location in the airport at time t_k, given Google has all my measurements from time $1:n$, $n > k$.

$$P(s_k \mid m_{1:n}) \propto P(s_k, m_{1:n}) = P(s_k, m_{1:k}, m_{k+1:n})$$

$$P(s_k, m_{1:n}) = P(m_{k+1:n} \mid s_k, m_{1:k})\, P(s_k, m_{1:k})$$

$P(ABC)$
$= P(C|AB)P(AB)$
$= P(C|AB)P(A|B)P(B)$

$$= P(m_{k+1:n} \mid s_k)\, P(s_k \mid m_{1:k})\, P(m_{1:k})$$

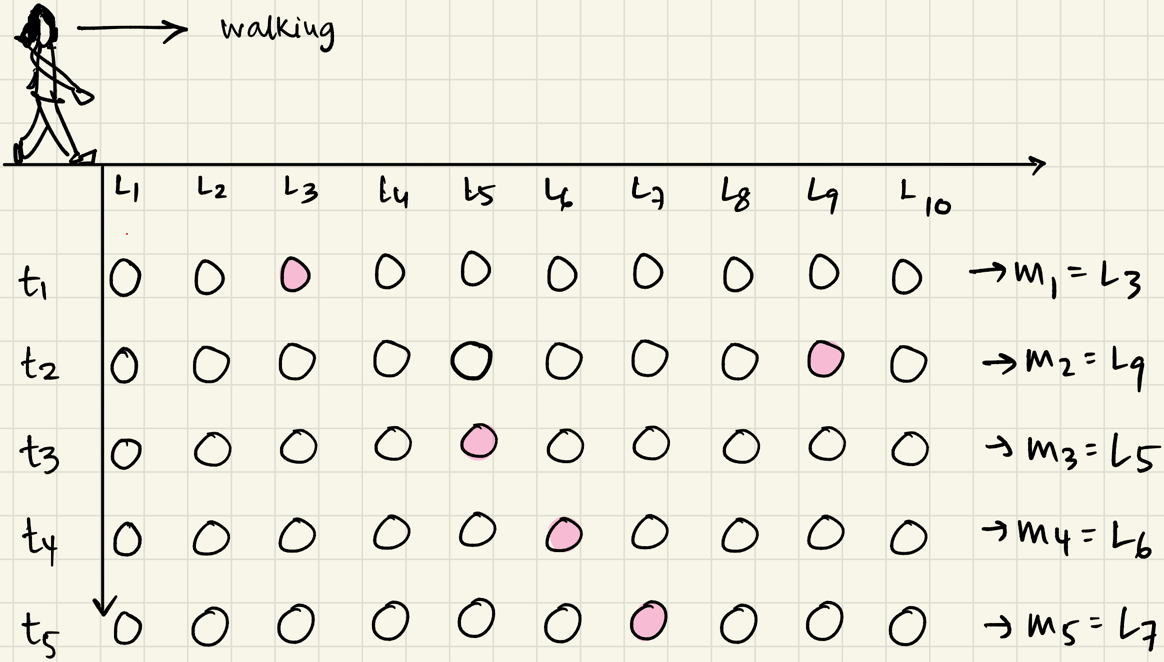
$$= P(s_k \mid m_{1:k}) \cdot P(m_{k+1:n} \mid s_k)$$

$P(s_k = UIUC \mid grades_{1:12})$

$P(s_k = UIUC \mid grade_{1:12})$ ← low

$P(grades_{13:17} \mid s_k = UIUC)$

$P(s_k \mid m_{1:k})$ ↓	$P(m_{k+1:n} \mid s_k)$ ↓
Forward (online, Real time)	**Backward (offline)**
↓	↓
Probability that murder suspect is at s_k = green st. given $k=4$ recent surveillance camera measurements of : main street → wright street → 6th street → university ave.	Probability that 4th to 8th measurement are Neil st. → Kirby road → Lincoln drive → university avenue, given suspect's k^{th} time location is s_k = green street.


walking
L1 L2 L3 L4 L5 L6 L7 L8 L9 L10
t1 → m1 = L3
t2 → m2 = L9
t3 → m3 = L5
t4 → m4 = L6
t5 → m5 = L7

→ Let's look at the **forward** component $P(s_k \mid m_{1:k})$

$$P(s_k \mid m_{1:k}) = \frac{P(s_k, m_{1:k})}{P(m_{1:k})} \propto P(s_k, m_{1:k})$$

$$s_k = \begin{bmatrix} L_1 \\ L_2 \\ \vdots \\ L_r \end{bmatrix}$$

By marginalizing

$$\text{RHS} = \sum_{s_{k-1}} P(s_k, s_{k-1}, m_{1:k})$$

$P(AB) = P(A|B)P(B)$

$$= \sum_{s_{k-1}} P(m_k \mid s_k, s_{k-1}, m_{1:k-1})\, P(s_k, s_{k-1}, m_{1:k-1})$$

$$= \sum_{s_{k-1}} P(m_k \mid s_k)\, P(s_k \mid s_{k-1}, m_{1:k-1})\, P(s_{k-1}, m_{1:k-1})$$

$$\underbrace{P(s_k, m_{1:k})}_{\alpha_k} = \sum_{s_{k-1}} P(m_k \mid s_k)\, P(s_k \mid s_{k-1})\, \underbrace{P(s_{k-1}, m_{1:k-1})}_{\alpha_{k-1}}$$

$$\therefore \quad \alpha_k = \sum_{s_{k-1}} P(m_k \mid s_k)\, P(s_k \mid s_{k-1})\, \alpha_{k-1}$$

Dynamic Programming

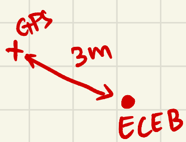
$$\alpha_1 = P(s_1, m_{1:1}) = P(s_1, m_1)$$

→ Initial condition $P(s_1, m_1) = P(m_1 \mid s_1)\, P(s_1)$ needs to be known.

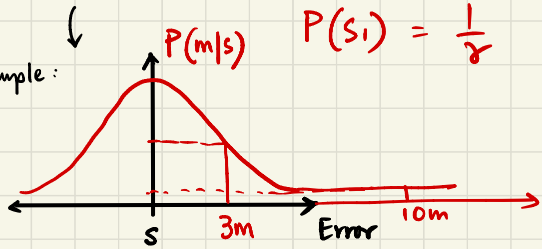
Errors of sensor from data sheet

perhaps all locations have equal probability

$$P(s_1) = \frac{1}{r}$$



Example:



→ Now let's look at the backward part : $P(m_{k+1:n} \mid S_k)$

$$P(m_{k+1:n} \mid S_k) = \frac{P(m_{k+1:n}, S_k)}{P(S_k)}$$

$$= \frac{1}{P(S_k)} \sum_{S_{k+1}} P(S_k, S_{k+1}, m_{k+1:n})$$

$$= \frac{1}{\cancel{P(S_k)}} \sum_{S_{k+1}} P(m_{k+2:n} \mid m_{k+1}, S_k, S_{k+1}) \cdot P(m_{k+1} \mid S_k, S_{k+1}) \cdot P(S_{k+1} \mid S_k)\, \cancel{P(S_k)}$$

$$P(m_{k+1:n} \mid S_k) = \sum_{S_{k+1}} P(m_{k+2:n} \mid S_{k+1})\, P(m_{k+1} \mid S_{k+1})\, P(S_{k+1} \mid S_k)$$

Say LHS $= P(m_{k+1:n} \mid S_k) = \beta_k$

$$\therefore \quad \beta_k = \sum_{S_{k+1}} \beta_{k+1}\, P(m_{k+1} \mid S_{k+1})\, P(S_{k+1} \mid S_k)$$

β_k, β_{k+1}: Dynamic Prog. again

$P(m_{k+1} \mid S_{k+1})$: Sensor error distr.

$P(S_{k+1} \mid S_k)$: Transition Prob. matrix

$$P(m_n \mid S_{n-1}) = \frac{P(m_n, S_{n-1})}{P(S_{n-1})} = \frac{1}{P(S_{n-1})} \sum_{S_n} P(S_n, S_{n-1}, m_n)$$

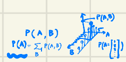
→ How should we initialize this β_k ?

$$\beta_{n-1} = P(m_{n:n} \mid S_{n-1}) = \sum_{S_n} \frac{P(m_n, S_{n-1}, S_n)}{P(S_{n-1})}$$

$$= \frac{1}{\cancel{P(S_{n-1})}} \sum_{S_n} P(m_n \mid S_n, S_{n-1})\, P(S_n \mid S_{n-1})\, \cancel{P(S_{n-1})}$$

$$\beta_{n-1} = \sum_{S_n} P(m_n \mid S_n)\, P(S_n \mid S_{n-1}) \Longrightarrow \text{Both are known}$$

$$\beta_{n-2} = \sum_{S_{n-1}} \beta_{n-1}\, P(m_{n-1} \mid S_{n-1})\, P(S_{n-1} \mid S_{n-2})$$



Recall original goal : $\boxed{P(s_k \mid m_{1:n})}$ ⇒ offline version

$$P(s_k \mid m_{1:n}) = \underbrace{P(s_k \mid m_{1:k})}_{\text{D.P}} \underbrace{P(m_{k+1:n} \mid s_k)}_{\text{D.P}}$$

$$\alpha_k = \sum_{s_{k-1}} P(m_k \mid s_k) P(s_k \mid s_{k-1}) \alpha_{k-1}$$

$$\beta_k = \sum_{s_{k+1}} \beta_{k+1} P(m_{k+1} \mid s_{k+1}) P(s_{k+1} \mid s_k)$$

HMM's ⇒ Efficiently identifying the most likely value of a state variable from a large space of possibilities.

⇒ Possible to also compute the full trajectory

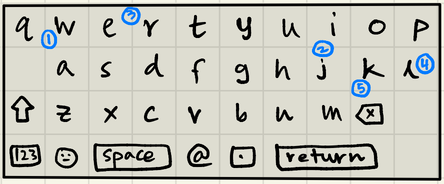
↳ called Viterbi Decoding

$P(s_{1:n} \mid m_{1:n})$

↳ Prob. of traj.

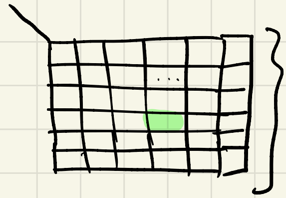
Some other applications of HMM (informal discussion)

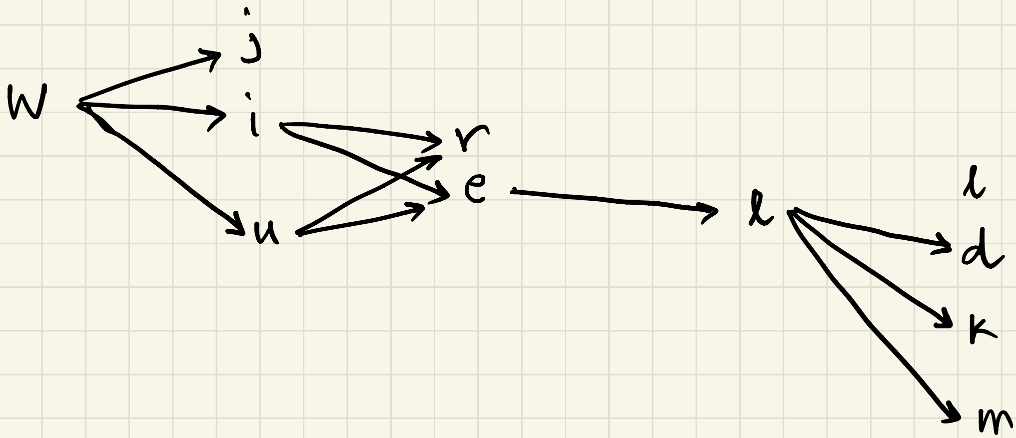
(I) in smartphone keyboard.



$P(m_1 | s_1) = P(\qquad) =$

$P(s_2 | s_1) = P(\qquad) =$





$\Downarrow$

Decodes to

Similar application in

$P(m_K | S_K) = P(\qquad)$