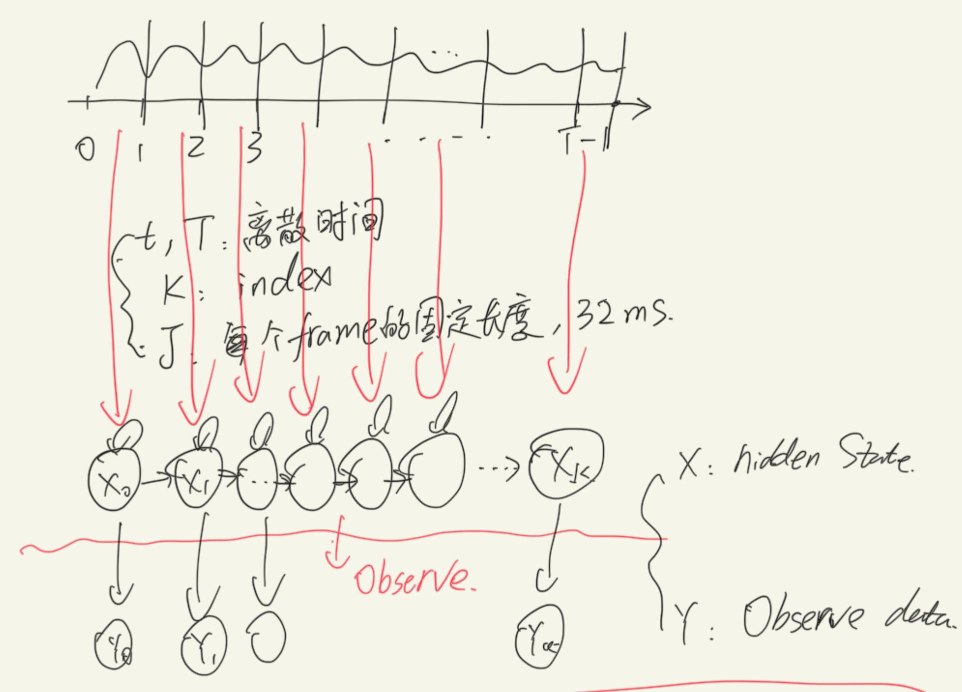


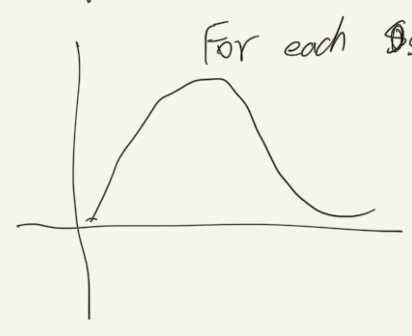
t, T: 离散时间
K: index
J: 每个frame的固定长度, 32 ms.

X: hidden State
Y: Observe data

目前不可用

posterior = $p(s|o) = \dfrac{p(o|s)\cdot p(s)}{p(o)}$

posterior distribution
For each $s \in \{S\}$, $\dfrac{p(o|s)p(s)}{p(o)}$ 可求



posterior = "已知观测O求状态S"

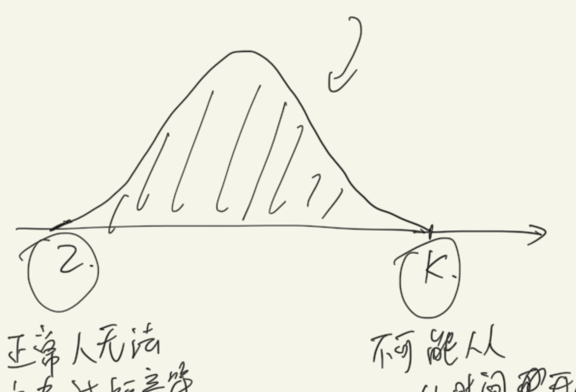
---

prior model: $p(s)$

state 是离散的 {rest, pitch, ⋯, artic, ⋯}
不同于 GMM 模型

$P_K = P(T > k+1 \mid T > k)$.

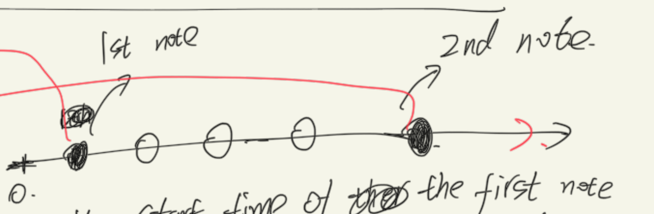
T: 花在某个note上面的时间(用帧计算).
32ms/帧



正常人无法演奏过短音符

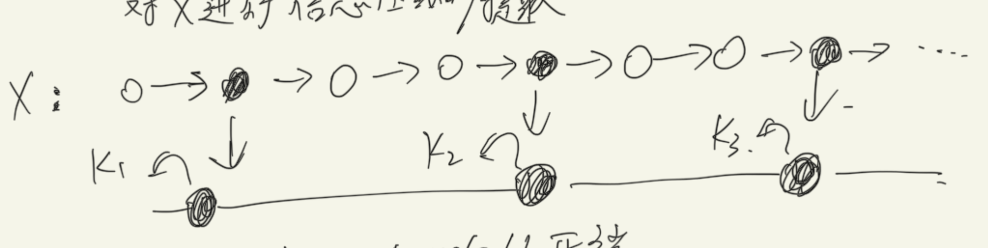
不可能从<0的时间开始演奏.

$K = \begin{pmatrix} K_1 \\ K_2 \\ K_3 \\ \vdots \\ K_m \end{pmatrix}$



K_1: start time of the first note
K_2: start time of the second note.
⋮
K_m.

可以把 Y → X 看成一个更精确的 vector.
对X进行信息压缩/提取



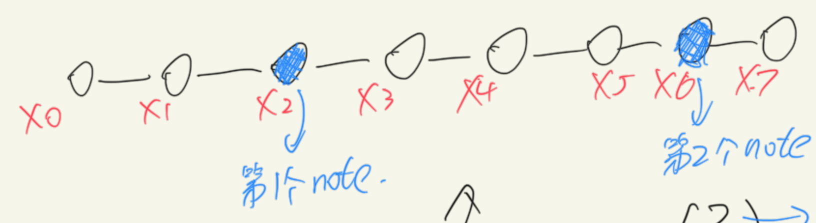
实心黑点: 一个 note 的开端.

$K(X) = (K_1(x), K_2(x) \cdots\cdots K_m(x))$.

$K_m(x)$ = the unique k of which $X_k = start(m)$.

m=1: $K_1(x)$ = the unique k letting $X_K = start(1)$ → m=1, 1: 第1个note.

⋮



所以, 给定 X = ↗, 得到 $K = \begin{pmatrix} 2 \\ 6 \end{pmatrix}$ → 第1个note对应的X的index: 2 → 第2个note对应的X的index: 6.

$P(K=k \mid Y=y) = \sum_{x:K(x)=k} P(X=x \mid Y=y)$.
→ why this works.?

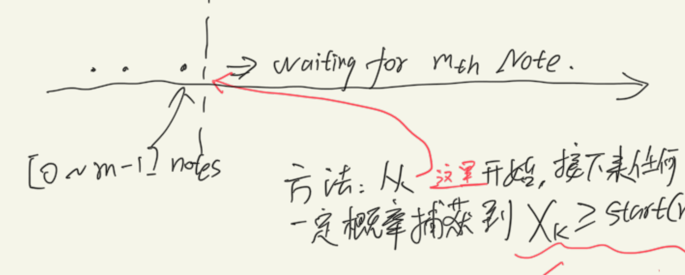
$P(|K_m - K_m(x)| \le \eta \mid Y=y) = \sum_{k=K_m-\eta}^{K_m+\eta} P(X_k = start(m) \mid Y=y)$ ⇒ 解释: 对固定的Y (观测结果)而言, 把所有[能够容忍η错误]的X的概率都加起来进行比较

$P(|K_1 - K_1(x)| \le \eta \mid Y=y) = \sum_{k=K_1-\eta}^{K_1+\eta} P(X_K = start(1) \mid Y=y)$

给定 Y=y, m=1时, 对 $K_1-\eta \xrightarrow{\eta\ K_1\ \eta} K_1+\eta$ 的所有 K,

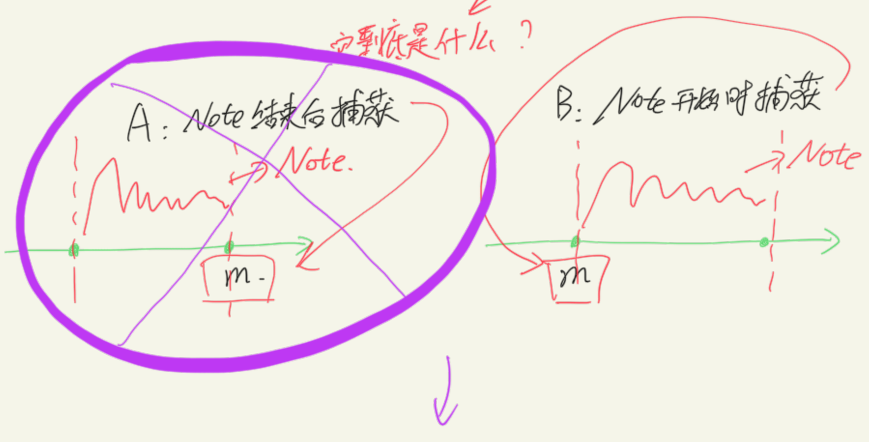
$\begin{cases} \text{音符从}(K_1-\eta)\text{开始 } P_1 \\ \text{音符从}(K_1)\text{开始 } P_2 \\ \text{音符从}(K_1+\eta)\text{开始 } P_3 \end{cases} \Rightarrow P_1+P_2+P_3$

Live accomplish.

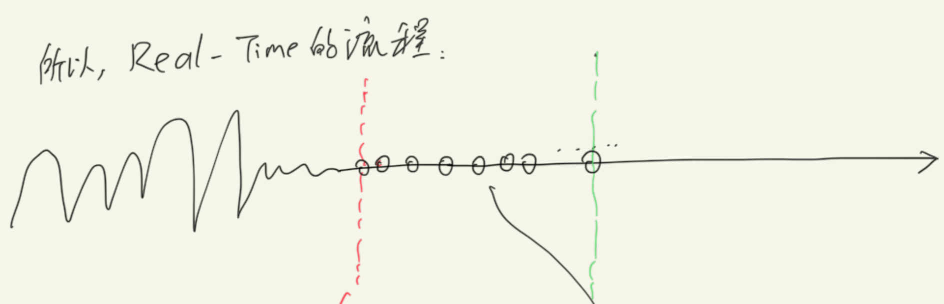


方法: 从这里开始, 接下来任何时刻都有一定概率捕获到 $X_K \ge start(m)$.

它到底是什么 ?



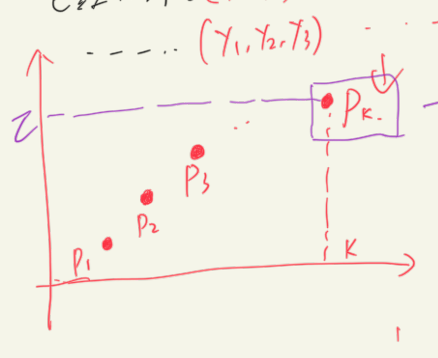
所以, Real-Time 的流程:



我们已经听到了 1, 2, 3 ⋯ (m-1) Note
开始监听第m个Note.

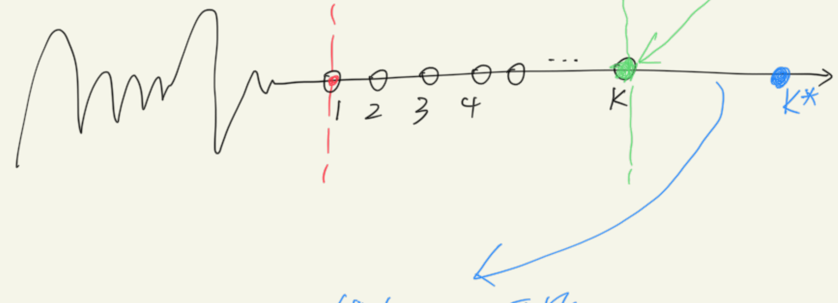
从红色虚线开始, 听到了 y_1, y_2, y_3 ⋯⋯

已经听到了 y_1, 那么第m个Note从X_1开始的几率有多大? P_1
已经听到了 (y_1, y_2), 那么第m个Note从X_2开始的几率有多大? P_2
⋯⋯ (y_1, y_2, y_3) ⋯⋯ X_3 ⋯⋯ P_3



→ 找到了!

找到了以后马上停止计算, 就认定第m个音符从 K开始



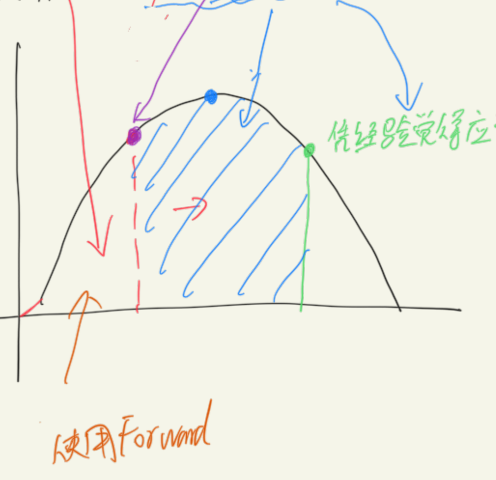
然后, 从 K开始:
从 K 到 K* 的范围内寻找最有可能的 start(m)

$\underset{K \le K^*}{\operatorname{argmax}} P(X_K = start(m) \mid Y_1 = y_1, Y_2 = y_2 \cdots Y_{K^*} = y_{K^*})$

注意这里:
条件是 "已经听完了 $y_1, y_2 \cdots y_{K^*}$"
而不是 $y_1 \to y_K$

Live 算法:

$\begin{cases} \text{从 } y_1 \text{ 开始, 直到找到第1个 } y_K\text{, 认为Note(m)从 } y_K \text{ 开始} \\ \text{从 } y_K \text{ 开始 过了一段时间 } K^*\text{, 从 } (y_K \to y_{K^*}) \text{ 之间找出更有可能的 } y_{argmax} \end{cases}$ Forward / Forward + Backward.



凭经验觉得应该差不多了

回头找最高点
(从蓝色区域)
使用Forward + Backward

不能仅仅使用Forward, 因为要观测完全部

使用Forward