7.1

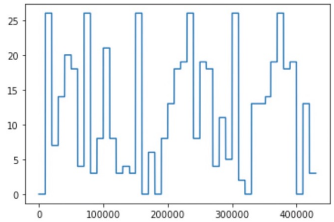
(a)

```python
import numpy as np

files =['transitionMatrix.txt','emissionMatrix.txt', 'initialStateDistribution.txt', 'observations.txt']

def parse_data(files):
    res = []
    for file in files:
        data = []
        f = open(file, 'r')
        for line in f:
            data.append([float(num) if '.' in num else int(num) for num in line.strip('\n').split()])
        f.close()
        res.append(np.array(data))
    return res


trans, emiss, pi, o = parse_data(files)
o = np.squeeze(o.reshape(-1, 1))

n = trans.shape[0] #27
m = emiss.shape[1] # 2
T = o.shape[0] # 430000
```

```python
l = np.empty((n, T))
for t in range(T):
    for i in range(n):
        if t == 0:
            l[i][0] = np.log(pi[i]) + np.log(emiss[i][o[0]])
        else:
            max_val = max([l[j][t-1] + np.log(trans[j][i]) for j in range(n)])
            l[i][t] = max_val + np.log(emiss[i][o[t]])

s = np.empty_like(o)
for t in range(T-1, -1, -1):
    curr_max = float('-inf')
    for i in range(n):
        if t == T - 1:
            if curr_max < l[i][t]:
                curr_max = l[i][t]
                s[t] = i
        else:
            if curr_max < (l[i][t] + np.log(trans[i][s[t+1]])):
                curr_max = (l[i][t] + np.log(trans[i][s[t+1]]))
                s[t] = i
```

(b)

In [37]:
```python
import matplotlib.pyplot as plt
plt.plot(s)
plt.show()
```



In [44]:
```python
string_long = [chr(ord('a')+ alphabet_num) if alphabet_num < 26 else ' ' for alphabet_num in s ]
string = string_long[0]
for i in range(1,len(string_long)):
    if string_long[i] != string_long[i-1]:
        string += string_long[i]
print(string)
```

```
a house divided against itself canot stand
```

7.2

(a) $P(S_{t+1}=j \mid S_t=i, o_1, o_2 \cdots o_T)$

$$= P(S_{t+1}=j \mid S_t=i, O_{t+1} \cdots O_T) \quad CI$$

$$= \frac{P(S_{t+1}=j, S_t=i, O_{t+1} \cdots O_T)}{P(S_t=i, O_{t+1} \cdots O_T)}$$

$$= \frac{\cancel{P(S_t=i)} \cdot P(S_{t+1}=j \mid S_t=i) \cdot P(O_{t+1} \cdots O_T \mid S_{t+1}=j)}{\cancel{P(S_t=i)} \cdot P(O_{t+1} \cdots O_T \mid S_t=i)}$$

CI

$$= \frac{a_{ij} \cdot P(O_{t+1} \mid \cancel{O_{t+2} \cdots O_T}, S_{t+1}=j) \cdot P(O_{t+2} \cdots O_T \mid S_{t+1}=j)}{\beta_{it}}$$

$$= \frac{a_{ij} \cdot b_j(O_{t+1}) \cdot \beta_{j(t+1)}}{\beta_{it}}$$

(b) $P(S_t = i \mid S_{t+1} = j, O_1, O_2 \cdots O_T)$

$$= P(S_t = i \mid S_{t+1} = j, O_1 \cdots O_t) \quad CI$$

$$= \frac{P(S_{t+1} = j \mid S_t = i, \overset{CI}{\cancel{O_1, O_2 \cdots O_t}}) \cdot \overset{\alpha_{it}}{P(S_t = i, O_1 \cdots O_t)}}{P(S_{t+1} = j, O_1 \cdots O_t)}$$

$$= \frac{P(S_{t+1} = j \mid S_t = i) \cdot P(S_t = i, O_1 \cdots O_t)}{\frac{P(S_{t+1} = j, O_1 \cdots O_{t+1})}{P(O_{t+1} \mid S_{t+1} = j, \overset{CI}{\cancel{O_1 \cdots O_t}})}}$$

$$= \frac{a_{ij} \cdot \alpha_{it} \cdot b_j(O_{t+1})}{\alpha_{j(t+1)}}$$

(C)

$$P(S_{t-1}=i, S_t=k, S_{t+1}=j \mid O_1, O_2 \cdots O_T)$$

$$= \frac{P(S_{t-1}=i, S_t=k, S_{t+1}=j, O_1, O_2 \cdots O_T)}{P(O_1 \cdots O_T)}$$

$$= \frac{\underbrace{\underbrace{P(O_1 \cdots O_{t-1}, S_{t-1}=i)}_{a} P(S_t=k \mid S_{t-1}=i \cancel{\cdots})}_{b} \overset{CI}{} P(O_t \mid S_t=k \cancel{\cdots}) P(S_{t+1}=j \mid S_t=k \overset{CI}{}) P(O_{t+1} \mid S_{t+1}=j \overset{CI}{\cdots}) P(O_{t+2} \cdots \mid S_{t+1}=j \overset{CI}{})}{\sum_i P(S_t=i, O_1 \cdots O_T)}$$

let x = # of hidden state value

$$= \frac{\alpha_i(t-1) \cdot a_{ik} \cdot b_k(O_t) \cdot a_{kj} \cdot b_j(O_{t+1}) \cdot \beta_j(t+1)}{\sum_x P(S_t=x, O_1 \cdots O_t) \cdot P(O_{t+1} \cdots O_T \mid S_t=x)}$$

$$= \frac{\alpha_i(t-1) \cdot a_{ik} \cdot b_k(O_t) \cdot a_{kj} \cdot b_j(O_{t+1}) \cdot \beta_j(t+1)}{\sum_x \alpha_{xt} \beta_{xt}}$$

(d) $P(S_{t+1}=j \mid S_{t-1}=i, O_1, O_2 \cdots O_T)$

$$= \frac{P(S_{t+1}=j, S_{t-1}=i, O_1 \cdots O_T)}{P(S_{t-1}=i, O_1 \cdots O_T)}$$

$$= \frac{\sum_x P(S_{t-1}=i, S_t=x, S_{t+1}=j, O_1 \cdots O_T)}{P(S_{t-1}=i, O_1 \cdots O_{t-1}) P(O_t \cdots O_T \mid S_{t-1}=i)} \quad ①$$

Part ① from c denominator, we have:

$$= \frac{\sum_x \cancel{\alpha_i(t-1)} a_{ix} \cdot b_x(O_t) \cdot a_{xj} \cdot b_j(O_{t+1}) \cdot \beta_j(t+1)}{\cancel{\alpha_i(t-1)} \beta_i(t-1)}$$

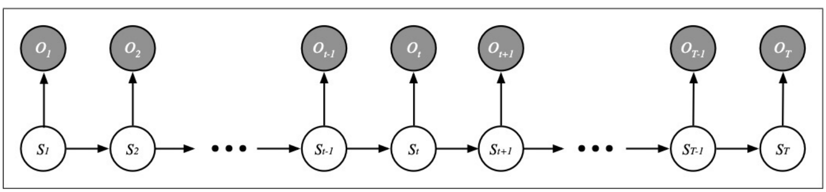
$$= \frac{\sum_x a_{ix} \cdot b_x(O_t) \cdot a_{xj} \cdot b_j(O_{t+1}) \cdot \beta_j(t+1)}{\beta_i(t-1)}$$

7.3

7.3 Conditional independence

Consider the hidden Markov model (HMM) shown below, with hidden states S_t and observations O_t for times $t \in \{1, 2, \ldots, T\}$. State whether the following statements of conditional independence are true or false.

F	$P(S_t \mid S_{t-1}) = P(S_t \mid S_{t-1}, S_{t+1})$
T	$P(S_t \mid S_{t-1}) = P(S_t \mid S_{t-1}, O_{t-1})$
F	$P(S_t \mid S_{t-1}) = P(S_t \mid S_{t-1}, O_t)$
F	$P(S_t \mid O_{t-1}) = P(S_t \mid O_1, O_2, \ldots, O_{t-1})$
T	$P(O_t \mid S_{t-1}) = P(O_t \mid S_{t-1}, O_{t-1})$
F	$P(O_t \mid O_{t-1}) = P(O_t \mid O_1, O_2, \ldots, O_{t-1})$
T	$P(S_2, S_3, \ldots, S_T \mid S_1) = \prod_{t=2}^{T} P(S_t \mid S_{t-1})$
T	$P(S_1, S_2, \ldots, S_{T-1} \mid S_T) = \prod_{t=1}^{T-1} P(S_t \mid S_{t+1})$
F	$P(S_1, S_2, \ldots, S_T \mid O_1, O_2, \ldots, O_T) = \prod_{t=1}^{T} P(S_t \mid O_t)$
F	$P(S_1, S_2, \ldots, S_T, O_1, O_2, \ldots, O_T) = \prod_{t=1}^{T} P(S_t, O_t)$
T	$P(O_1, O_2, \ldots, O_T \mid S_1, S_2, \ldots, S_T) = \prod_{t=1}^{T} P(O_t \mid S_t)$
T	$P(O_1, O_2, \ldots, O_T) = \prod_{t=1}^{T} P(O_t \mid O_1, \ldots, O_{t-1})$



7.4

(a) $q_{jt} = P(S_t=j \mid o_1 \cdots o_t)$

$$= \frac{P(S_t=j \mid o_1 \cdots o_{t-1})\, P(o_t \mid S_t=j, \overset{CI}{\cancel{o_1 \cdots o_{t-1}}})}{P(o_t \mid o_1 \cdots o_{t-1})}$$

$P(a,b \mid c) = P(b \mid a c) P(a \mid c)$

$$= \frac{\sum_i P(\overset{a}{S_{t-1}=i}, \overset{b}{S_t=j} \overset{|}{\mid} \overset{c}{o_1 \cdots o_{t-1}})\, P(o_t \mid S_t=j)}{P(o_t \mid o_1 \cdots o_{t-1})}$$

$$= \frac{\sum_i P(S_t=j \mid S_{t-1}=i, \overset{CI}{\cancel{o_1 \cdots o_{t-1}}}) \cdot P(S_{t-1}=i \mid o_1 \cdots o_{t-1})\, P(o_t \mid S_t=j)}{\sum_{ij} P(S_{t-1}=i, S_t=j, o_t \mid o_1 \cdots o_{t-1})}$$

$$= \frac{\sum_i P(S_t=j \mid S_{t-1}=i, \overset{CI}{\cancel{o_1 \cdots o_{t-1}}}) \cdot P(S_{t-1}=i \mid o_1 \cdots o_{t-1})\, P(o_t \mid S_t=j)}{\sum_{ij} P(S_t=j, o_t \mid S_{t-1}=i, \underset{CI}{\cancel{o_1 \cdots o_{t-1}}}) \cdot P(S_{t-1}=i \mid o_1 \cdots o_{t-1})}$$

$$= \frac{\sum_i P(S_t=j \mid S_{t-1}=i, \overset{CI}{\cancel{o_1 \cdots o_{t-1}}}) \cdot P(S_{t-1}=i \mid o_1 \cdots o_{t-1})\, P(o_t \mid S_t=j)}{\sum_{i,j} P(o_t \mid S_t=j, \underset{CI}{\cancel{S_{t-1}=i}}) \cdot P(S_t=j \mid S_{t-1}=i) P(S_{t-1}=i \mid o_1 \cdots o_{t-1})}$$

$$= \frac{\sum_i a_{ij}\, q_{i(t-1)}\, b_j(o_t)}{\sum_{i,j} b_j(o_t)\, a_{ij} \cdot q_{i(t-1)}} = \frac{b_j(o_t) \sum_i a_{ij} q_{i(t-1)}}{Z_t}$$

(b) The joint distribution of hidden state and observation in a continuous setting will have to be multivariate Gaussian to compute the intergral, which maynot be the case