Q1. $P(\pi_3 = \text{'R'} \mid X) = \dfrac{P(x_1 x_2 x_3 \wedge \pi_3 = \text{'R'}) \cdot P(x_4 x_5 \mid \pi_3 = R)}{P(X)} = \dfrac{f_R(3) \cdot g_R(3)}{f_R(5) + f_V(5)}$

$f_k(i+1) = e_k(x_{i+1}) \sum_{j \in \pi} f_j(i) \cdot t_{jk}$ $\quad g_k(i) = \sum t_{kj}\, e_j(x_{i+1})\, g_j(i+1)$

e	A	C	G	T
k=V	.25	.25	.25	.25
k=R	.91	.03	.03	.03

t_{12}	V	R
V	.75	.25
R	.10	.90

BOUNDARIES

$f_B(0) = 1 \quad P(\pi_1 = \text{'R'}) = \frac{1}{2}$

$g_R(n) = 1 \quad g_V(n) = 1 \quad (n=5)$

π_3 'R'

X = A T A A A
x_1 x_2 x_3 x_4 x_5

$f_k(i)$

i	'R'	'V'
1	.455	.125
2	0.01322	0.03481
3	0.01875	$6.857 \cdot 10^{-3}$
4	0.01692	$1.754 \cdot 10^{-3}$
5	0.01426	$7.519 \cdot 10^{-4}$

$g_k(i)$

i	R	V
1		
2		
3	0.7016	0.2698
4	0.844	0.415
5	E=1	E=1

$P(\pi_3 = \text{'R'} \mid X) = \dfrac{f_R(3) \cdot g_R(3)}{f_R(5) + f_V(5)} = \dfrac{0.01875 \cdot 0.7016}{(0.01426 + 7.519 \cdot 10^{-4})}$

$\boxed{= 0.8763}$

$f_R(1) = e_R(A) \cdot f_B(0) \cdot t_{BR} = .91 \cdot 1 \cdot \frac{1}{2} = .455$

$f_V(1) = e_V(A) \cdot f_B(0) \cdot t_{BV} = .25 \cdot 1 \cdot \frac{1}{2} = .125$

CALCULATOR...

rounding

$\downarrow$

$f_R(2) = e_R(T) \cdot [f_R(1) \cdot t_{RR} + f_V(1) \cdot t_{VR}] = .03 \cdot [.455 \cdot .90 + .125 \cdot .25] = 0.0132$

$f_V(2) = e_V(T) \cdot [f_R(1) \cdot t_{RV} + f_V(1) \cdot t_{VV}] = .25 \cdot [.455 \cdot .10 + .125 \cdot .75] = 0.0348$

$f_R(3) = e_R(A) \cdot [f_R(2) \cdot t_{RR} + f_V(2) \cdot t_{VR}] = .91 \cdot [.01322 \cdot .90 + .03481 \cdot .25] = .01875$

$f_V(3) = e_V(A) \cdot [f_R(2) \cdot t_{RV} + f_V(2) \cdot t_{VV}] = .25 \cdot [.01322 \cdot .10 + .03481 \cdot .75] = 6.857e^{-3}$

$f_R(4) = .91 \cdot [.01875 \cdot .90 + 6.857e^{-3} \cdot .25] = 0.01692$

$f_V(4) = .25 \cdot [.01875 \cdot .10 + 6.857e^{-3} \cdot .75] = 1.754 \cdot 10^{-3}$

$f_R(5) = .91 \cdot [.01692 \cdot .90 + 1.754e^{-3} \cdot .25] = 0.01426$

$\cancel{f_V(5) = .25 [1.754e^{-3} \cdot .10 +}$

$f_V(5) = .25 [.01692 \cdot .10 + 1.754e^{-3} \cdot .75] = 7.519 \cdot 10^{-4}$

REVERSE

$\cancel{g_R(4) = t_{RR} e_R(A) g_R(5) \cdot .90} \cdot$

$g_R(4) = t_{RR} e_R(A) g_R(5) + t_{RV} e_V(A) g_V(5) = .90 \cdot .91 \cdot 1 + .10 \cdot .25 \cdot 1 = 0.844$

$g_V(4) = t_{VR} e_R(A) g_R(5) + t_{VV} e_V(A) g_V(5) = .25 \cdot .91 \cdot 1 + .75 \cdot .25 \cdot 1 = 0.415$

$g_R(3) = t_{RR} e_R(A) g_R(4) + t_{RV} e_V(A) g_V(4) = .90 \cdot .91 \cdot 0.844 + .10 \cdot .25 \cdot .415 = 0.7016$

$g_V(3) = t_{VR} e_R(A) g_R(4) + t_{VV} e_V(A) g_V(4) = .25 \cdot .91 \cdot 0.844 + .75 \cdot .25 \cdot .415 = 0.2698$

Q2: 24, 33, 42, 30, 44, 38, 27, 39, 47, 51

For Binomial Distribution:

$E(x) = np$

$P(x=k|\theta) = \binom{n}{k} \theta^k (1-\theta)^{n-k}$

WE COULD ALSO USE 2nd MOMENT: $\sigma_x^2 = np(1-p)$ BUT WE WON'T...

1. Method of Moments:

$\mu_x = np$

$\theta = \frac{\mu_x}{n} = \frac{(24+33+42+30+44+38+27+39+47+51)}{10} \cdot \frac{1}{1000} = \boxed{.0375}$

2. Least Squares

$SS = (24-n\theta)^2 + (33-n\theta)^2 + (42-n\theta)^2 + \dots + (51-n\theta)^2$

$\frac{dSS}{d\theta} = -2n(24-n\theta) - 2n(33-n\theta) - 2n(42-n\theta) + \dots - 2n(51-n\theta) = 0$ $\div -2n$

$= 24 - n\theta + 33 - n\theta + \dots + 51 - n\theta \Rightarrow \theta = \frac{24+33+\dots+51}{10n} = \frac{375}{10000} = \boxed{.0375}$

3. Maximum Likelihood

X $P(D|\theta) = \theta^{24}(1-\theta)^{976} \cdot \theta^{33}(1-\theta)^{967} \dots$

$= \theta^{375}(1-\theta)^{9625}$ (oops, need $\binom{1000}{24}$ combinations: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

$P(D|\theta) = \frac{1000!}{24!\,976!} \cdot \theta^{24}(1-\theta)^{976} \cdot \frac{1000!}{33!\,967!} \theta^{33}(1-\theta)^{967}$

$P(D|\theta) = \left(\text{SUPER MESSY CONSTANT}\right) \cdot \theta^{375}(1-\theta)^{9625}$

$\ln[P(D|\theta)] = \ln C + 375 \ln\theta + 9625 \ln(1-\theta)$

COMBINATION DISAPPEARS WITH DERIVATIVE

$\frac{d}{d\theta} \ln[L(x)] = \cancel{C} + \frac{375}{\theta} + \frac{-9625}{1-\theta}$

SET $\frac{dL}{d\theta} = 0$

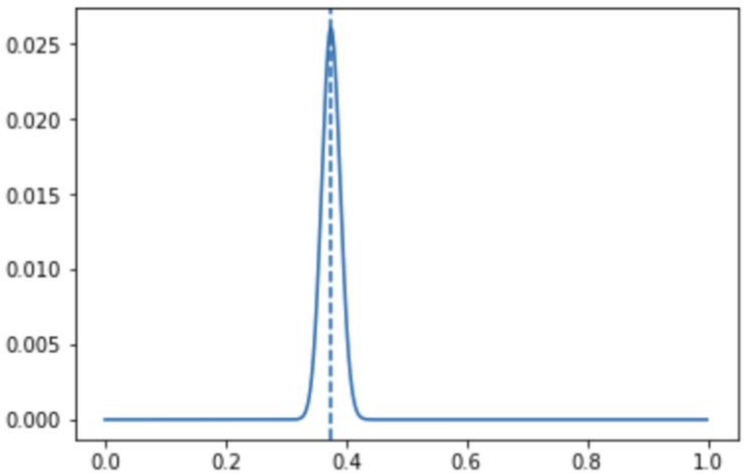
$= \frac{375(1-\theta)}{\theta(1-\theta)} + \frac{-9625\theta}{\theta(1-\theta)} = \frac{375 - 375\theta - 9625\theta}{\theta(1-\theta)} = 0$

$-10000\theta = -375$

$\theta = \frac{375}{10000} = \boxed{.0375}$

Bayesian method probability distribution:

```
0.375
```




```
#! /usr/bin/env python

## Code for generating plot above and MAP estimate:

import numpy as np
from matplotlib import pyplot as plt

def theta_likelihood(thet):
    likelihood = np.power(thet,375) * np.power(1-thet,625)
    return likelihood

thet_values = np.arange(0,1,.001)
pdf_values = np.zeros_like(thet_values)

for n in range(len(thet_values)):
    thet = thet_values[n]
    pdf_values[n] = theta_likelihood(thet)

pdf_values = pdf_values / sum(pdf_values)
map_estimate = thet_values[np.argmax(pdf_values)]
print(map_estimate)

#print(pdf_values)
plt.plot(thet_values,pdf_values)
plt.axvline(map_estimate,linestyle='dashed')
plt.show()
```

4. BAYESIAN METHOD

ASSUME UNIFORM DISTRIBUTION FOR $P(\theta)$!

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{\int P(D|\theta)P(\theta)d\theta}$$

$$= \frac{1}{\int P(D|\theta)P(\theta)d\theta} \cdot P(D|\theta)$$

LET $C = \frac{1}{\int P(D|\theta)\cancel{P(\theta)}d\theta}$

$$= C \cdot P(D|\theta)$$

$$= C_2 \cdot \theta^{375}(1-\theta)^{9625}$$

LET $C_2 = \frac{\text{MESSY COMBINATORIALS}}{\int P(D|\theta)\cancel{P(\theta)}d\theta)}$

$P(\theta) = 1$ EVERYWHERE

Q3.

S1 - AGCTA

S2 - ACCGA

$k = 2$

$T_1 =$

	A	C	G	T
A	.85	.05	.05	.05
C	.05	.85	.05	.05
G	.05	.05	.85	.05
T	.05	.05	.05	.85

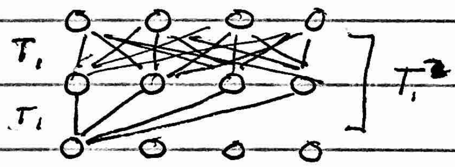
ANCESTRAL PROBS

$P(A) = P(C) = P(G) = P(T) = 0.25$

← PROBABILITIES FOR SUBSTITUTION AFTER 1 TRANSITION (GENERATION)

1: $P(A \cap A | A_a) \cdot P(A_a) + P(A \cap A | C_a) \cdot P(C_a) + P(A \cap A | G_a) \cdot P(G_a) + P(A \cap A | T_a) \cdot P(T_a)$

$P(A \cap A | A_a) = P(A_{S1} | A_a) \cdot P(A_{S2} | A_a)$



$P(A_2 | A_a) = P(A_2 | A_1) \cdot P(A_1) + P(A_2 | C_1) \cdot P(C_1) \ldots$

$P(A_1) = P(A_1 | A_a) P(A_a) + P(A_1 | C_a) P(C_a) \ldots$

RATHER THAN DO RECURSION, WE CAN USE MATRIX MULTIPLICATION →

$$T^2 = \begin{bmatrix} .73 & .09 & .09 & .09 \\ .09 & .73 & .09 & .09 \\ .09 & .09 & .73 & .09 \\ .09 & .09 & .09 & .73 \end{bmatrix}$$

$P(A \cap A | \mathbf{Q}) = .73 \cdot .73 \cdot .25 + .09 \cdot .09 \cdot .25 + .09 \cdot .09 \cdot .25 + .09 \cdot .09 \cdot .25$

$A \cap A$ $\quad pos\ 1 = .73 \cdot .73 \cdot .25 + 3(.09 \cdot .09 \cdot .25) = 0.1393$

$G \cap C$ $\quad pos\ 2 = P(G \cap C \mid A_a) P(A_a) + \dots = (.09 \cdot .09 \cdot .25) \cdot 2 + (.73 \cdot .09 \cdot .25) \cdot 2 = 0.036$

$pos\ 3 = pos\ 1 = 0.1393$

$pos\ 4 = pos\ 2 = 0.0369$

$pos\ 5 = pos\ 1 = 0.1393$

$$PROB(S_1 \cap S_2 \mid a) = pos1 \cdot pos2 \cdot pos3 \cdot pos4 \cdot pos5$$
$$= 0.1393 \cdot 0.0369 \cdot 0.1393 \cdot 0.0369 \cdot 0.1393$$
$$= 3.68 \cdot 10^{-6}$$

Q3.

Calculating T2, probabilities:

In [15]:
```
import numpy as np

T = np.ones([4,4]) / 20
for n in range(4):
    T[n,n] = .85

print(T) # Gen 1, looking great
print(T*T) # That's not right...we need dot product
print(np.dot(T,T)) # That looks better.
```

```
[[0.85 0.05 0.05 0.05]
 [0.05 0.85 0.05 0.05]
 [0.05 0.05 0.85 0.05]
 [0.05 0.05 0.05 0.85]]
[[0.7225 0.0025 0.0025 0.0025]
 [0.0025 0.7225 0.0025 0.0025]
 [0.0025 0.0025 0.7225 0.0025]
 [0.0025 0.0025 0.0025 0.7225]]
[[0.73 0.09 0.09 0.09]
 [0.09 0.73 0.09 0.09]
 [0.09 0.09 0.73 0.09]
 [0.09 0.09 0.09 0.73]]
```

In [14]:
```
pos1 = .73 * .73 * .25 + 3 * (.09 * .09 * .25)
pos2 = 2 * (.09 * .09 * .25) + 2 * (.73 * .09 * .25)

print(pos1 * pos2 * pos1 * pos2 * pos1)
```

Out[14]: 3.680493724705767e-06