3.1)

a) $A_{ij} = P(X_{T+1} = j \mid X_T = i)$

$$P(X_{T+1} \mid X_1) = \sum_{m_T}^{m} P(X_{T+1}, X_T = m_T \mid X_1) \qquad \text{marg.}$$

$$= \underbrace{\sum_{m_2} \sum_{m_3} \dots \sum_{m_T}}_{T-1} P(X_{T+1}, X_T, \dots X_2 \mid X_1)$$

$$= \underbrace{\sum_{m_2} \sum_{m_3} \dots \sum_{m_T}}_{T-1} P(X_{T+1} \mid X_T, X_{T-1}, \dots X_1) \cdot P(X_T \mid X_{T-1}, X_{T-2} \dots X_1) \cdot$$

$$\vdots$$

$$P(X_2 \mid X_1)$$

$$= \overbrace{\sum_{m_2} \sum_{m_3} \dots \sum_{m_T}}^{T-1} P(X_{T+1} \mid X_T) P(X_T \mid X_{T-1}) \dots P(X_2 \mid X_1)$$

$$= \overbrace{\sum_{m_2} \sum_{m_3} \dots \sum_{m_T}}^{T-1} A_{i m_T} A_{m_T m_{T-1}} \dots\dots A_{m_2 j} \quad \dots (i)$$

For matrix multiplication:-

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} \qquad \text{for 2 matrices}$$

If they're the same matrices

$$\Rightarrow b_{ij} = \sum_{k=1}^{n} a_{ik} a_{kj} \qquad [B = A^2_{n \times n}]$$

Similarly, for 3 matrices

$$b_{ij} = \sum_{k_1}^{n} \sum_{k_2}^{n} a_{ik_1} a_{k_1 k_2} a_{k_2 j} \qquad [B = A^3]$$

This, can be generalized for m matrices:-

$$b_{ij} = \sum_{k_1}^{n} \sum_{k_2}^{n} \cdots \sum_{k_{m-1}}^{n} a_{ik_1} a_{k_1 k_2} \cdots\cdots a_{k_{m-1} j}$$

$$[B = A^m]$$

Thus, (i) equates to

$$\sum_{m_2} \sum_{m_3} \cdots \sum_{m_T} A_{i m_T} A_{m_T m_{T-1}} \cdots\cdots A_{m_2 j} = [A^t_{ij}]$$

$$\Rightarrow \boxed{P(X_{t+1} = j \mid X_1 = i) = [A^t_{ij}]}$$

Using induction:-

- Base case :- $t = 1$

$$P(X_{1+1} = j \mid X_1 = i) = [A^1]_{ij}$$

$$\Rightarrow \quad P(X_2 = j \mid X_1 = i) = A_{ij}$$

Which is true, since we're given

$$P(X_{t+1} = j \mid X_t = i) = A_{ij}$$

- Assuming, the statement is true for 't-1'

$$\Rightarrow \quad P(X_t = j \mid X_1 = i) = [A^{t-1}]_{ij}$$

$\Rightarrow$ • For 't' we have:-

$$P(X_{t+1} = j \mid X_1 = i) = \sum_k^m P(X_{t+1} = j, X_t = k \mid X_1 = i)$$

marg.

$$= \sum_k^m P(X_{t+1} = j \mid X_t = k, X_1 = i) P(X_t = k \mid X_1 = i)$$

P.R.

$$= \sum_k P(X_{t+1} = j \mid X_t = k) P(X_t = k \mid X_1 = i)$$

d-separation case 1

$$= \sum_k A^{t-1}_{ik} A_{kj}$$

$$\boxed{P(X_{t+1} = j \mid X_1 = i) = [A^t]_{ij}}$$

b) $P(X_2 | X_1) = A_{ij}$

$P(X_3 | X_1) = \sum_m A_{im} A_{mj}$

For calculating $[A^2]_{ij}$ we will require m iterations.
But, to calculate $[A^2]_{im}$ $\forall m \in \{1 \ldots m\}$

we will repeat this process 'm' times

$\Rightarrow$ Asymptotic complexity $= O(m^2)$

$P(X_4 | X_1) = \sum_m [A^2]_{im} A_{mj}$ $\qquad O(m^2)$

$$\left.\begin{array}{cc} \vdots & \vdots \\ \vdots & \vdots \end{array}\right\} t-1 \text{ times}$$

$P(X_{T+1} | X_1) = \sum_m [A^{t-1}]_{im} A_{mj}$

$\Rightarrow$ Total asymptotic complexity $= O(m^2 t)$

c/

Since,

$$P(X_{t+1} = j \mid X_1 = i) = [A^t]_{ij}$$

Now, similar to the concept of binary numbers, we can represent any number t as a sum of 2^i.

eg

$$4 = 100 = 1 \times 2^2 + 0 \times 2^1 + 0 \times 2^0$$

thus, in general if t_{10} & t_2 represent decimal and binary representation of 't'

$$\Rightarrow \cancel{t_{10} = \sum_{i=0}^{\log t_{10}} t_2[i]}$$

$$\Rightarrow t_{10} = \sum_{i=0}^{\log_2 t_{10}} [t_2]_i \times 2^i$$

where, $[t_2]_i$ represents i^{th} least significant bit

$\Rightarrow$ We can split the matrix multiplication A^t into $\log_2 t$ parts :-

$$A^{t_{10}} = \prod_{i=0}^{\log_2 t} A^{[t_2]_i 2^i}$$

$\Rightarrow$ Since complexity of matrix multiplication is $O(m^3)$

$\Rightarrow$ Total asymptotic complexity $= O(m^3 \log_2 t)$

d) $P(X_t = j \mid X_1 = i) = \sum_k P(X_t = j, X_{t+1} = k \mid X_1 = i)$ (marg)

$= \sum_k P(X_{t+1} = k \mid X_t = j) P(X_t = j \mid X_1 = i)$ (P.R)

$= \sum_k A_{jk} [A^{t-1}]_{ij}$ (From (a))

$\Rightarrow$ Each element in the row of A needs to be multiplied by each column of A^{t-1}, but since $s \ll m$

$\Rightarrow$ Only s rows will strictly be used.

~~$\Rightarrow$ Asympt~~

$\Rightarrow$ Complexity of this step will be $O(sm)$

$\because$ this is repeated $t-1$ times

$\Rightarrow$ Asymptotic complexity of the calculation is $O(smt)$

e) $P(X_1 = i \mid X_{t+1} = j) =$

$$\frac{P(X_{t+1} = j \mid X_1 = i)\, P(X_1 = i)}{P(X_{t+1} = j)}$$ (Bayes)

$$\Rightarrow \frac{P(X_{t+1} = j \mid X_1 = i)\, P(X_1 = i)}{\sum_{k_1} \cdots \sum_{k_t} P(X_{t+1}, X_t, \ldots X_1)}$$

$$\Rightarrow \frac{[A^t]_{ij}\, P(X_1 = i)}{\sum_{k_1} \cdots \sum_{k_t} P(X_{t+1} \mid X_t)\, P(X_t \mid X_{t-1}) \cdots P(X_2 \mid X_1)\, P(X_1)}$$

$$\Rightarrow \frac{[A^t]_{ij}\, P(X_1 = i)}{\sum_{k_1} [A^t]_{k_1 j}\, P(X_1 = k_1)}$$

$$\Rightarrow \boxed{P(X_1 = i \mid X_{t+1} = j) = \frac{[A^t]_{ij}\, P(X_1 = i)}{\sum_{k} [A^t]_{kj}\, P(X_1 = k)}}$$

3.2

a) $P(Y_1 = y_1 | X_1 = x_1)$

$$= \sum_{x_2} P(Y_1, X_2 = x_2 | X_1) \quad \boxed{\text{marg.}}$$

$$= \sum_{x_2} P(Y_1 | X_1, X_2 = x_2) P(X_2 = x_2 | X_1) \quad \boxed{\text{P.R.}}$$

$$= \sum_{x_2} P(Y_1 | X_1, X_2 = x_2) P(X_2 = x_2) \quad \boxed{\text{marginal independence}}$$

b) $$P(Y_1) = \sum_{x_1} \sum_{x_2} P(Y_1, X_1 = x_1, X_2 = x_2) \quad \boxed{\text{marg}}$$

$$= \sum_{x_1} \sum_{x_2} P(Y_1 | X_1 = x_1, X_2 = x_2) P(X_2 = x_2 | X_1 = x_1) \times P(X_1 = x_1) \quad \boxed{\text{P.R.}}$$

$$= \sum_{x_1} \sum_{x} P(Y_1 | x_1, x_2) P(x_2) P(x_1) \quad \boxed{\text{marginal independence}}$$

c) $P(X_n | Y_1, Y_2, \ldots Y_{n-1})$

Given the evidence set $E = \phi$ and using Y_n as the sink node for rule-3 of d-separation

$$P(X_n | Y_1, Y_2, \ldots Y_{n-1}) = P(X_n)$$

d) $P(Y_n | X_n, Y_1, Y_2 \ldots Y_{n-1})$

$$\Rightarrow \sum_{x_{n-1}} P(Y_n, x_{n-1} | X_n, Y_1, Y_2 \ldots Y_{n-1})$$

$$\Rightarrow \sum_{x_{n-1}} P(Y_n | X_n, x_{n-1}, Y_1, Y_2 \ldots Y_{n-1})$$
$$P(X_{n-1} | Y_1, Y_2 \ldots Y_{n-1}, X_n)$$

~~$P(X_{n-1} | X_n | Y_1$~~

$$P(X_{n-1} | X_n, Y_1, Y_2 \ldots Y_{n-1})$$
$$= P(X_{n-1} | Y_1, Y_2 \ldots Y_{n-1})$$ rule-3 d-separation

$$P(Y_n | X_n, X_{n-1}, Y_1, Y_2 \ldots Y_{n-1})$$
$$= P(Y_n | X_n, X_{n-1}, Y_1, Y_2 \ldots Y_{n-1})$$

rule - 2 d-separation

$$\Rightarrow \sum_{x_{n-1}} P(Y_n \mid X_n, X_{n-1}, Y_1 \ldots Y_{n-1}) P(X_{n-1} \mid Y_1 \ldots Y_{n-1}, X_n)$$

$$= \boxed{\sum_{x_{n-1}} P(Y_n \mid X_n, X_{n-1}) P(X_{n-1} \mid Y_1 \ldots Y_{n-1})}$$

e) $P(Y_n \mid Y_1, \ldots Y_{n-1})$

$$= \sum_{x_{n-1}} P(Y_n, X_{n-1} \mid Y_1, \ldots Y_{n-1}) \quad \boxed{\text{marg}}$$

$$= \sum_{x_{n-1}} P(Y_n \mid X_{n-1}, Y_1, \ldots Y_{n-1}) P(X_{n-1} \mid Y_1, \ldots Y_{n-1}) \quad \boxed{\text{P.R}}$$

$$= \sum_{x_{n-1}} P(Y_n \mid X_{n-1}) P(X_{n-1} \mid Y_1 \ldots Y_{n-1}) \quad \boxed{\begin{array}{c}\text{Rule-2}\\ \text{d-separation}\end{array}}$$

$$= \sum_{x_{n-1}} \sum_{x_n} P(Y_n, X_n \mid X_{n-1}) P(X_{n-1} \mid Y_1 \ldots Y_{n-1})$$

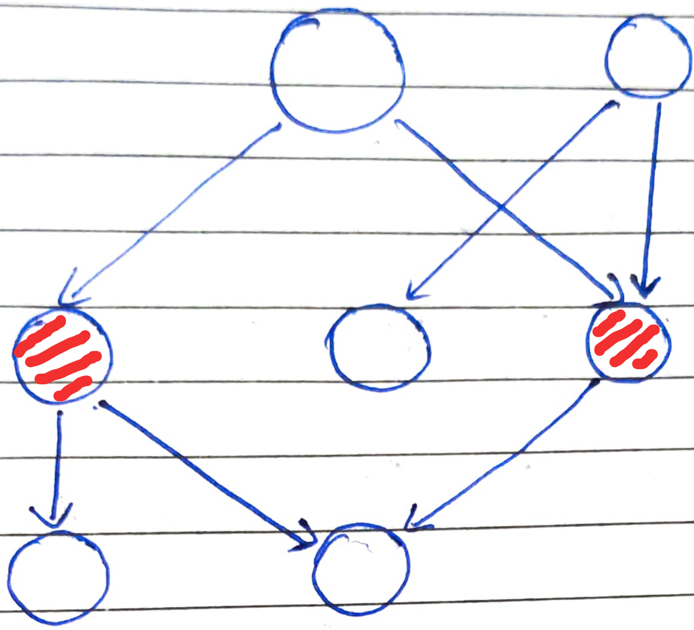
$$= \sum_{x_{n-1}} \sum_{x_n} P(Y_n \mid X_{n-1}, X_n) P(X_n \mid X_{n-1}) \cdot P(X_{n-1} \mid Y_1 \ldots Y_{n-1})$$

$$= \boxed{\sum_{x_{n-1}} \sum_{n} P(Y_n \mid X_{n-1}, X_n) P(X_n) P(X_{n-1} \mid Y_1, \ldots Y_{n-1})}$$

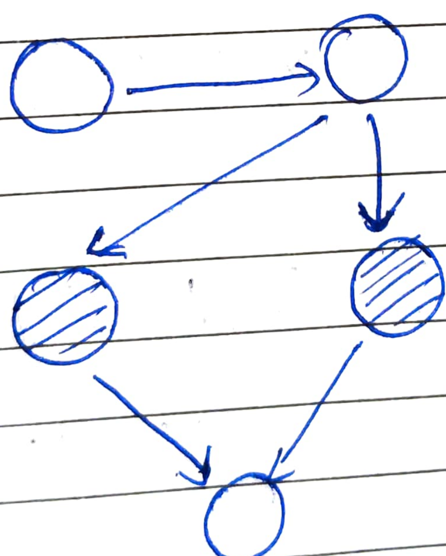
3.3

a) Polytree

b)



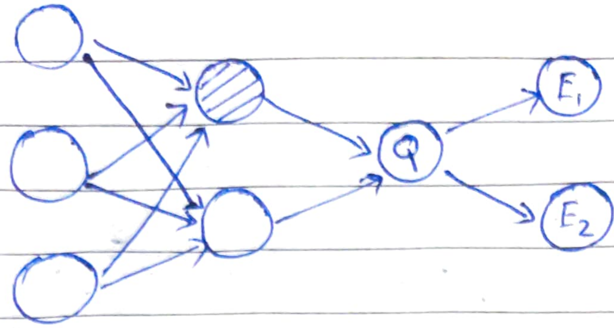
c)



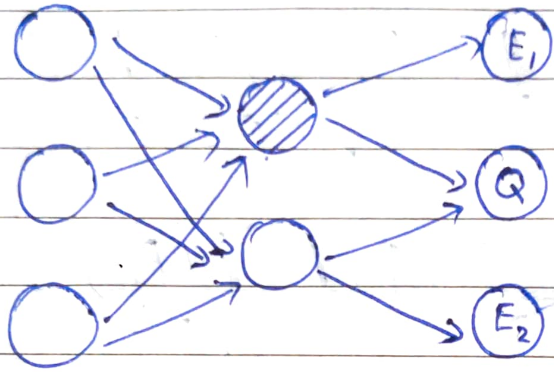
d) Polytree

e) Polytree

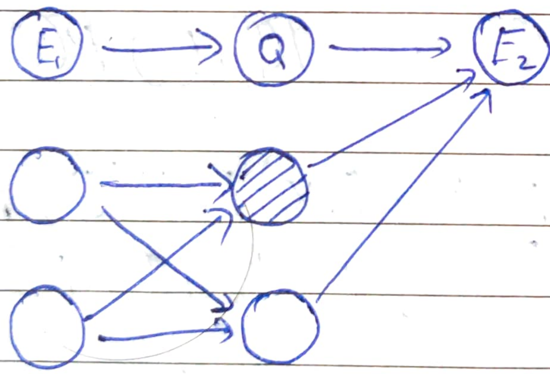
3.4



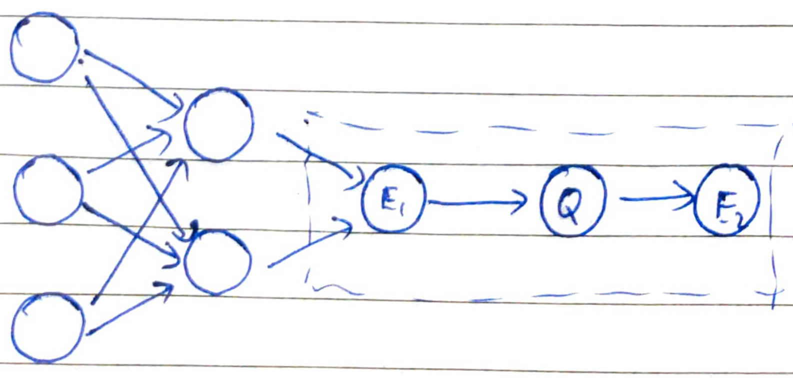
Q's parents belong to loopy network.



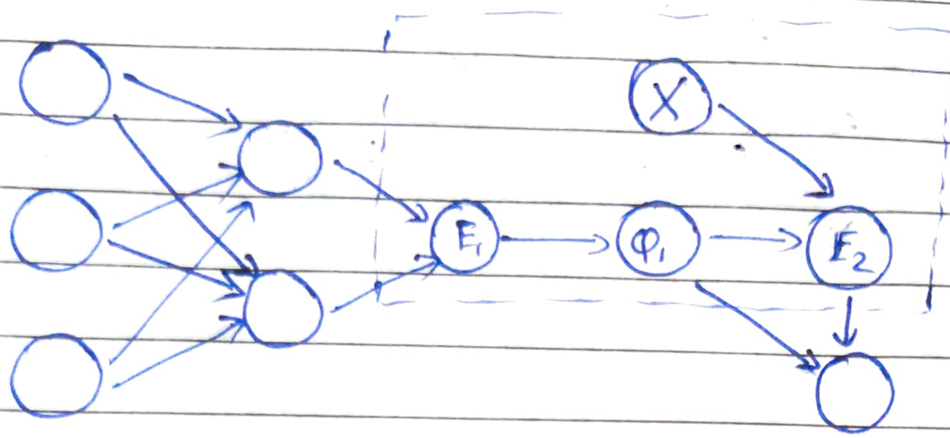
Q's parents belong to loopy network.



Q's siblings belong in a loopy network



E_1 & E_2 are Q's markov blanket.



$$P(\varphi \mid E_1, E_2) = \frac{P(E_2 \mid \varphi, E_1)\, P(\varphi \mid E_1)}{P(E_2 \mid E_1)}$$

$$= \frac{P(E_2 \mid \varphi)\, P(\varphi \mid E_1)}{\sum_{q} P(E_2, \varphi \mid E_1)}$$

$$= \frac{\sum_{x} P(E_2, x \mid \varphi)\, P(\varphi \mid E_1)}{\sum_{q} P(E_2 \mid \varphi, E_1)\, P(\varphi \mid E_1)}$$

$$= \sum_{x} \frac{P(E_2 \mid \varphi, x)\, P(x \mid \varphi)\, P(\varphi \mid E_1)}{\sum_{q} P(E_2 \mid \varphi)\, P(\varphi \mid E_1)}$$

$$= \sum_{x} \frac{P(E_2 \mid \varphi, x)\, P(x)\, P(\varphi \mid E_1)}{\sum_{q} \sum_{x} P(E_2 \mid \varphi, x)\, P(x)\, P(\varphi \mid E_1)}$$

Thus, we can calculate the probability without using children of φ_1 & E_2

3.5

Y	$P(Y\mid X=0)$	$P(Y\mid X=1)$
1	0.09375	0.09375
2	0.28125	0.09375
3	0.09375	0.03125
4	0.03125	0.28125
5	0.28125	0.03125
6	0.09375	0.28125
7	0.03125	0.09375
8	0.09375	0.09375

$$P(Y\mid X) = P(Y_1\mid X)\,P(Y_2\mid X)\,P(Y_3\mid X)$$

Y	$P(Z_1=1\mid Y)$	$P(Z_2=1\mid Y)$
1	0.9	0.1
2	0.8	0.2
3	0.7	0.3
4	0.6	0.4
5	0.5	0.5
6	0.4	0.6
7	0.3	0.7
8	0.2	0.8

$$P(Z\mid Y) = P(Z\mid Y_1, Y_2, Y_3)$$

3.6 a) $P(Z|B_1, B_2, B_3 \ldots B_n) = \left(\frac{1-\alpha}{1+\alpha}\right)\alpha^{|z-f(B)|}$

$$\sum_z P(Z|B_1, \ldots B_n) = \sum_z \left(\frac{1-\alpha}{1+\alpha}\right)\alpha^{|z-f(B)|}$$

$\forall z \in [-\infty, \infty]$

$$\Rightarrow \left(\frac{1-\alpha}{1+\alpha}\right)\sum_z \alpha^{|z-f(B)|}$$

let $k = |z - f(B)| \Rightarrow k = 0$ for $z = f(B)$

$$\Rightarrow \left(\frac{1-\alpha}{1+\alpha}\right)\left(1 + 2\sum_k \alpha^k\right) \quad \forall k \in [1, \infty]$$

$\because 0 < \alpha < 1$

The sum of the geometric series is

$$\sum_k^{\infty} \alpha^k = \frac{\alpha}{1-\alpha}$$

$$\Rightarrow \left(\frac{1-\alpha}{1+\alpha}\right)\left(1 + \frac{2 \times \alpha}{1-\alpha}\right) = \left(\frac{1-\alpha}{1+\alpha}\right)\left(\frac{1+\alpha}{1-\alpha}\right) = 1$$

$$\Rightarrow \boxed{\sum_z P(Z|B_1, \ldots B_n) = 1}$$

3.6

In [2]:

```
import random
import matplotlib.pyplot as plt
```

In [3]:

```
def P_B_i():
    num = random.random()
    if num < 0.5:
        return 0
    else:
        return 1
```

In [4]:

```
def f_B(B, n):
    num = 0
    for i in range(n):
        num = num + (B[i] * pow(2, i))
    return num
```

In [5]:

```
def P_Z_B(Z, B, n, alpha):
    result1 = pow(alpha, abs(Z - f_B(B, n)))
    const = ((1 - alpha)/(1 + alpha))
    result = const * result1
    return result
```

In [6]:

```
def indicator(x, x_):
    if x == x_:
        return 1
    return 0
```

In [1]:

```
def likelihood_estimation(targetBit, iterations):
    numerator = 0
    denominator = 0
    estimate = []
    for i in range(iterations):
        for bit in range(n):
            B[bit] = P_B_i()

        likelihood_weight = P_Z_B(Z, B, n, alpha)
        numerator = numerator + (indicator(B[targetBit], 1) * likelihood_weight)
        denominator = denominator + likelihood_weight
        estimate.append(numerator/denominator)
    return estimate
```

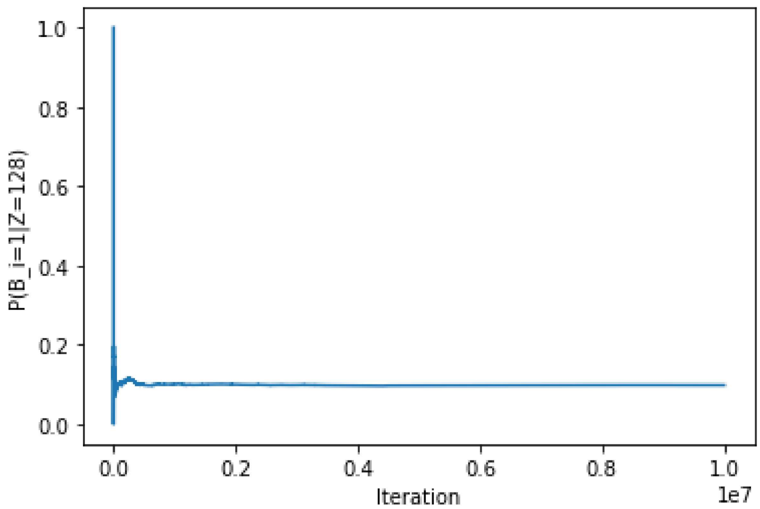
In [23]:

```
n = 10
B = [0 for i in range(n)]
Z = 128
alpha = 0.1
iterations = 10000000
```

i = 2

In [33]:

```
estimates = likelihood_estimation(1, iterations)
plt.plot(estimates)
plt.ylabel('P(B_{}=1|Z=128)'.format(2))
plt.xlabel('Iteration')
plt.show()
```



In [ ]:

```

```

In [34]:

```
estimates[-1]
```

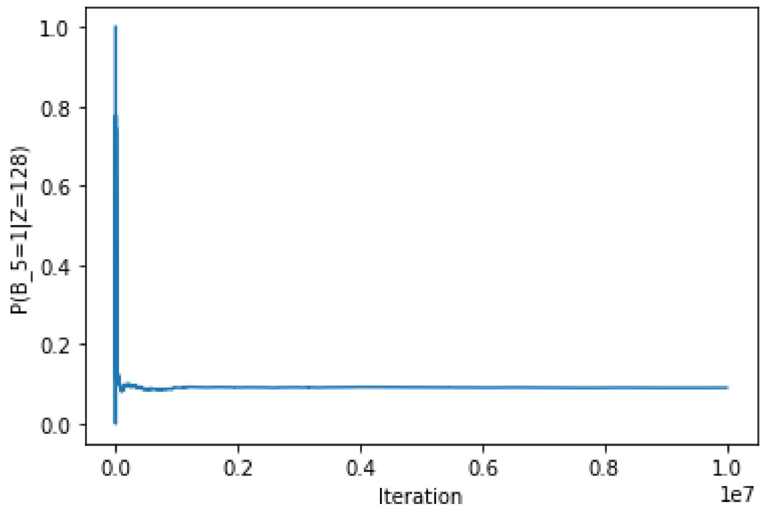
Out[34]:

```
0.09816876389177609
```

i = 5

In [40]:

```
estimates = likelihood_estimation(4, iterations)
plt.plot(estimates)
plt.ylabel('P(B_{}=1|Z=128)'.format(5))
plt.xlabel('Iteration')
plt.show()
```



In [41]:

```
estimates[-1]
```

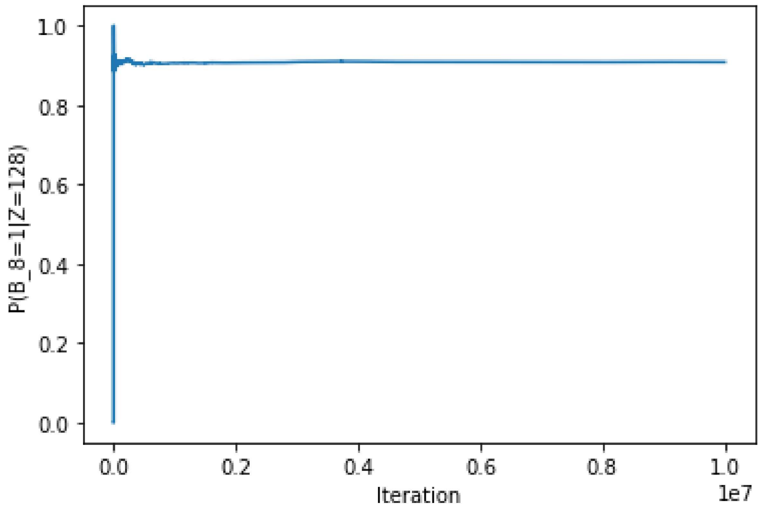
Out[41]:

```
0.09019225688900386
```

i = 8

In [43]:

```
estimates = likelihood_estimation(7, iterations)
plt.plot(estimates)
plt.ylabel('P(B_{}=1|Z=128)'.format(8))
plt.xlabel('Iteration')
plt.show()
```



In [44]:

```
estimates[-1]
```

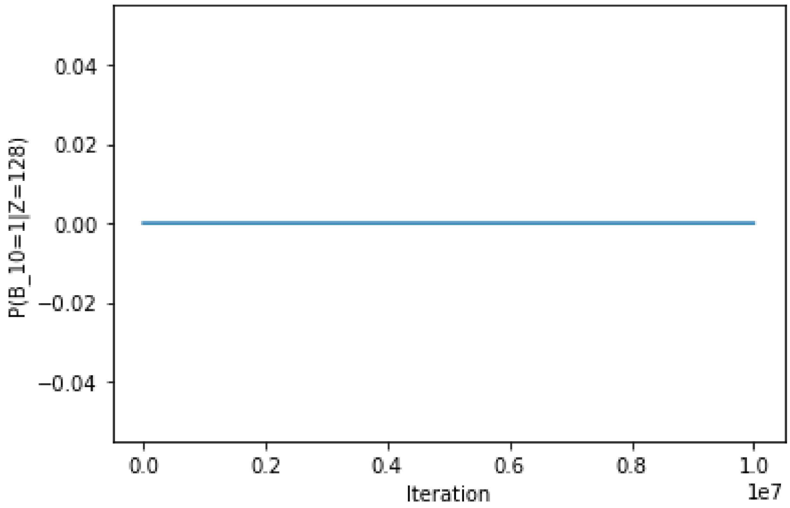
Out[44]:

0.9082078987581559

i = 10

In [50]:

```
estimates = likelihood_estimation(9, iterations)
plt.plot(estimates)
plt.ylabel('P(B_{}=1|Z=128)'.format(10))
plt.xlabel('Iteration')
plt.show()
```



In [51]:

```
estimates[-1]
```

Out[51]:

```
0.0
```

3.7

a) $P(B|A,C,D)$

$$= \frac{P(D|A,B,C)\,P(B|A,C)}{P(D|A,C)}$$ Bayes rule

∵ We know everything with respect to D instead of B, thus we use bayes rule to infer $P(D|Pa(D))$ instead

$$= \frac{P(D|B,C)\,P(B|A)}{P(D|A,C)}$$ rule-1 & rule 3 of D-separation

$$= \frac{P(D|B,C)\,P(B|A)}{\sum_b P(D,B|A,C)}$$ marginalization

Since we want to introduce independence between D & A using B

$$= \frac{P(D|B,C)\,P(B|A)}{\sum_b P(D|A,B,C)\,P(B|A,C)}$$ Product Rule

$$= \frac{P(D|B,C)\,P(B|A)}{\sum_b P(D|B,C)\,P(B|A)}$$ rule-1 & rule-3 of d-separation

b) $P(B \mid A, C, D, E, F)$

$= P(B \mid A, C, D, E)$ rule-2 d-separation

$= P(B \mid A, C, D)$ rule-3 of d-separation.

This is the same calculation that we did in part (a)

$$\Rightarrow \boxed{P(B \mid A, C, D, E, F) = \frac{P(D \mid B C)\, P(B \mid A)}{\sum_{b} P(D \mid B C)\, P(B \mid A)}}$$

c) $P(B, E, F \mid A, C, D)$

$= P(B \mid A, C, D, E, F)\, P(E \mid A, C, D, F)\, P(F \mid A, C, D)$

$\boxed{P.R}$

$= P(B \mid A, C, D)\, P(E \mid A, C, D, F)\, P(F \mid A, C, D)$

From (b)

$= P(B \mid A, C, D)\, P(E \mid C)\, P(F \mid A)$

rule-2 of d-separation.

$$\Rightarrow \boxed{P(B, E, F \mid A, C, D) = P(B \mid A, C, D)\, P(E \mid C)\, P(F \mid A)}$$

3.8

a) $$P(Q=q \mid E=e) \simeq \frac{\sum_{i=1}^{T} I(q,q_i)\,P(E=e \mid Y=y_i, Z=z_i)}{\sum_{i=1}^{T} P(E=e \mid Y=y_i, Z=z_i)}$$

b) $$P(Q_1=q_1, Q_2=q_2 \mid E_1=e_1, E_2=e_2)$$

$$\simeq \frac{\sum_{i}^{T} I(q_1,q_{1i})\,I(q_2,q_{2i})\,P(E_1=e_1 \mid Q_1=q_{1i}, X=x_i) \times P(E_2=e_2 \mid E_1=e_1, Z=z_i)}{\sum_{i=1}^{T} P(E_1=e_1 \mid Q=q_i, X=x_i)\,P(E_2=e_2 \mid E_1=e_i, Z=z_i)}$$