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

$$= \frac{P(a, b, c, d)}{P(c, d)}$$

$$= \frac{P(a)P(b \mid a)P(c \mid a, b)P(d \mid b, c)}{\sum_a \sum_b P(a, b, c, d)}$$

$$= \frac{P(a)P(b \mid a)P(c \mid a, b)P(d \mid b, c)}{\sum_a \sum_b P(a)P(b \mid a)P(c \mid a, b)P(d \mid b, c)}$$

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

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

$P(b \mid c, d)$

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

(c) $$\mathcal{L} = \sum_{t}^{T} \log P(C=c_t, D=d_t)$$

$$= \sum_{t}^{T} \log \sum_{A} \sum_{B} P(A=a, B=b, c=c_t, d=d_t)$$

$$= \sum_{t}^{T} \log \sum_{A} \sum_{B} P(A=a)\, P(B=b \mid A=a) P(C=c_t \mid A=a, B=b) P(D=d_t \mid B=b, C=c_t)$$

$$= \sum_{t}^{T} \log \sum_{a} \sum_{b} P(a) P(b \mid a) P(c_t \mid a, b) P(d_t \mid b, c_t)$$

(d)

$$P(A=a) \leftarrow \frac{1}{T} \sum_{t}^{T} P(a \mid c_t, d_t)$$

$$P(B=b \mid A=a) \leftarrow \sum_{t}^{T} \frac{P(a, b \mid c_t, d_t)}{P(a \mid c_t, d_t)}$$

$$P(C=c \mid A=a, B=b) \leftarrow \sum_{t}^{T} \frac{P(a, b \mid c_t, d_t) I(c, c_t)}{P(a, b \mid c_t, d_t)}$$

$$P(D=d \mid B=b, C=c) \leftarrow \sum_{t}^{T} \frac{P(b \mid c_t, d_t) I(c, c_t)(d, d_t)}{P(b \mid c_t, d_t) I(c, c_t)}$$

(a)

$$P(Y=1 \mid X) = \sum_{Z \in \{0,1\}^n} P(Y=1, Z \mid X)$$

$$= \sum_{Z \in \{0,1\}^n} P(Y=1, Z) \cdot P(Z \mid X)$$

$$= \sum_{Z \in \{0,1\}^n} (1 - I(Z, \{0\}^n)) P(Z \mid X)$$

$$= \sum_{Z \in \{0,1\}^n} P(Z \mid X) - P(Z \in \{0\}^n \mid X)$$

$$= 1 - \prod_{i}^{n} P(Z_i = 0 \mid X_i)$$

$$= 1 - \prod_{i}^{n} (1 - p_i)^{X_i}$$

(b)

$$P(z_i=1, x_i=1 \mid x, y)$$

$$= I(x_i=1)\, P(z_i=1 \mid x, y)$$

$$= I(x_i=1)\left(\frac{P(y \mid x, z_i=1)\, P(z_i=1 \mid x)}{P(y \mid x)}\right)$$

$$= I(x_i=1)\left(\frac{I(y=1)\, P(z_i=1 \mid x_i)}{P(y \mid x)}\right)$$

$$= I(x_i=1)\left(\frac{I(y=1)\quad I(x_i=1)\, p_i}{P(y \mid x)}\right)$$

$$= I(x_i=1)\left(\frac{I(y=1)\, p_i}{P(y \mid x)}\right)$$

$\because$ $P(y \mid x)$ in denominator, so only do $P(y=1 \mid x)$

$$\Rightarrow \frac{x_i y_i p_i}{P(y=1 \mid x)}$$

$$= \frac{x_i y_i p_i}{1 - \prod_j (1-p_j)^{x_j}}$$

(C) $P(z_i|X_i) \leftarrow \dfrac{\sum_t^T P(z_i, x_i | X, y)}{\sum_t^T P(X_i | X, y)}$

Only non-zero X, therefore non-zero Z is going to be contributing to $P(z_i|X_i)$ [denomintor non-zero]

$$\Rightarrow P(Z_i=1 | X_i=1) \to \frac{\sum_t^T P(z_i=1, x_i=1 | X, Y)}{\sum_t^T I(X_i=1, x)}$$

$$P(Z_i=1 | X_i=1) \to \frac{\sum_t^T P(z_i=1, x_i=1 | X=x^t, Y=y^t)}{T_i}$$

$$p_i \to \frac{\sum_t^T P(z_i=1, x_i=1 | X=x^t, Y=y^t)}{T_i}$$

(d) $L = \frac{1}{T}\sum_{t}^{T} \log P(Y = \vec{y}^t \mid X = \vec{x}^t)$

$= \frac{1}{T}\sum_{t}^{T} (1-y_t)\log P(y=0 \mid \vec{x}_t) + y_t \log P(y=1 \mid \vec{x}_t)$

$= \frac{1}{T}\sum_{t}^{T} (1-y_t)\log x_{it}(1-p_i) + y_t \log\left(1 - \prod_{i=1}^{n}(1-p_i)^{x_{it}}\right)$

In [267]:
```python
import numpy as np
import math
```

In [268]:
```python
files =['X.txt', 'Y.txt']
def parse_data(files):
    res = []
    for file in files:
        data = []
        f = open(file, 'r')
        for line in f:
            data.append([int(num) for num in line.strip('\n').split()])
        f.close()
        res.append(data)
    return res


X, Y = parse_data(files)

X = np.array(X)
Y = np.array([y[0] for y in Y])
p = np.ones((23,1)) * 0.05
```

In [269]:
```python

def log_likelihood(p,X,Y):
    return 1 / X.shape[0] * np.sum(Y[:,np.newaxis] * np.log(1-np.exp(X.dot(np.log(1-p)))) + (1-Y[:,np.newaxis]) * X.dot
```

In [270]:
```python
def M_update(p,X,Y):
    before = (X.T * p).T * Y[:,np.newaxis] / (1 - np.exp(X.dot(np.log(1-p))))
    new = (np.sum(before,axis = 0) / np.sum(X,axis = 0))[:,np.newaxis]
    return new
```

In [271]:
```python
def predict(p,X,Y):
    y_prob = 1 - np.exp(X.dot(np.log(1-p)))
    y_pred = np.where(y_prob >= 0.5,1,0)
    return np.sum(np.abs(y_pred - Y[:,np.newaxis]))
```

In [272]:
```python
likelihood_list = [log_likelihood(p,X,Y)]
error_list = [predict(p,X,Y)]
print(f"i={0}   mistiakes={error_list[0]}  likelihood={likelihood_list[0]}")
max_i = 256
for i in range(1,max_i+1):
    p = M_update(p,X,Y)
    error = predict(p,X,Y)
    error_list.append(error)
    likelihood = log_likelihood(p,X,Y)
    likelihood_list.append(likelihood)
    if math.log(i, 2).is_integer():
print(f"i={i}   mistiakes={error}  likelihood={likelihood}")
```

```
i=0   mistiakes=175  likelihood=-0.958085408215791
i=1   mistiakes=56  likelihood=-0.49591639407753624
i=2   mistiakes=43  likelihood=-0.4082208170583913
i=4   mistiakes=42  likelihood=-0.3646149825001877
i=8   mistiakes=44  likelihood=-0.3475006162087825
i=16   mistiakes=40  likelihood=-0.33461704895854844
i=32   mistiakes=37  likelihood=-0.3225814031674977
i=64   mistiakes=37  likelihood=-0.31482669836285615
i=128   mistiakes=36  likelihood=-0.3111558472151894
i=256   mistiakes=36  likelihood=-0.31016135347407603
```

6.4

(a) $f(x) = \log\cosh(x)$

$$f'(x) = \frac{\sinh(x)}{\cosh(x)} = \tanh(x)$$

$$f'(x) = 0 \rightarrow x = 0$$

$$f''(x) = \frac{d}{dx}\tanh(x) = \frac{1}{\cosh^2 x} = \operatorname{sech}^2 x$$

$$f''(x=0) > 0$$

$\therefore$ $x=0$ is the minimum

(b) $f''(x) = \operatorname{sech}^2 x$

$$= \left(\frac{2}{e^x + e^{-x}}\right)^2$$

By definition, $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$

[Taylor series], so $e^x \geq 1+x$, $e^{-x} \geq 1-x$

$$f''(x) = \left(\frac{2}{e^x + e^{-x}}\right)^2$$

$$\leq \left(\frac{2}{1+x+1-x}\right)^2$$

$$\leq 1$$

C.

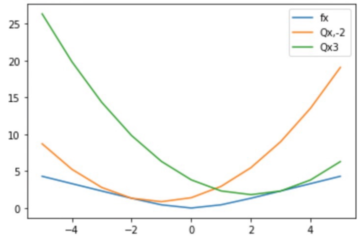
In [24]:
```
import matplotlib.pyplot as plt
import numpy as np
import math

x1 = [i for i in range(-5, 6)]
y1 = [math.log(math.cosh(i)) for i in x1]

y2 = [math.log(math.cosh(-2))+math.tanh(-2)*(i+2)+ 0.5*(i+2)**2 for i in x1]
y3 = [math.log(math.cosh(3))+math.tanh(3)*(i-3)+ 0.5*(i-3)**2 for i in x1]

plt.plot(x1, y1, label="fx")
plt.plot(x1, y2, label="Qx,-2")
plt.plot(x1, y3, label="Qx3")
plt.legend()

# show the plot
plt.show()
```



(d) $Q(x,y) = f(y) + f'(y)(x-y) + \frac{1}{2}(x-y)^2$

$$Q(x,x) = f(x) + f'(x)(x-x) + \frac{1}{2}(x-x)^2$$
$$= f(x)$$

$$f(x) = f(y) + \int_y^x f'(u)\,du$$
$$= f(y) + \left(\int_y^x du \left(f'(y) + \int_y^u f''(v)\,dv\right)\right)$$

from (b) $f''x \leq 1$

$$\Rightarrow f(x) \leq f(y) + f'(y)(x-y) + \int_y^x du \int_y^u 1\,dv$$
$$\leq f(y) + f'(y)(x-y) + \frac{1}{2}(x-y)^2$$
$$= Q(x,y)$$

e)

$$\frac{d}{dx}Q(x,x_n) = \frac{d}{dx}\left(f(x_n) + f'(x_n)(x-x_n) + \frac{1}{2}(x-x_n)^2\right)$$

$$= \frac{d}{dx}\left(\cancel{f(x_n)} + f'(x_n)x - \cancel{f'(x_n)\cdot x_n} + \frac{1}{2}x^2 - xx_n + \cancel{\frac{x_n^2}{2}}\right)$$

$$= f'(x_n) + x - x_n$$

$$\frac{d}{dx}Q = 0 \Rightarrow f'(x_n) + x - x_n = 0$$

$$\Rightarrow x = x_n - f'(x_n)$$

$$\frac{d^2}{dx^2}Q(x,x_n) = \frac{d}{dx}\left(f'(x_n) + x - x_n\right)$$

$$= 1$$

when $x = x_n - f'(x_n)$, $\frac{d^2}{dx^2}Q > 0$

$\therefore$ argmin $Q(x,x_n)$ when $x = x_n - f'(x_n)$

$$x_{n+1} = x_n - f'(x_n) = x_n - \tanh(x_n)$$