MLT: Week-4

EM Algorithm

A Aniruddha

E.M Algorithm

E.M Algorithm

Initialize θ^0 = $\left\{\begin{matrix} \mu_1^0, & ... & , \mu_k^0 \\ {\sigma^2}_1^0, & ... & , {\sigma^2}_k^0 \\ \pi_1^0, & ... & , \pi_k^0 \end{matrix}\right\}$

E.M Algorithm

Initialize θ^0 = $\begin{Bmatrix} \mu_1^0, & ... & ,\mu_k^0 \\ {\sigma^2}_1^0, & ... & ,{\sigma^2}_k^0 \\ \pi_1^0, & ... & ,\pi_k^0 \end{Bmatrix}$

Until Convergence $(||\theta^{t+1} - \theta^t|| \leq \epsilon)$, where ϵ is the tolerance parameter, do the following:

$$\lambda^{t+1} = \arg\max_{\lambda} \ \texttt{modified\_log}(\theta^t, \lambda) \qquad \rightarrow \text{Expectation Step}$$

$$\theta^{t+1} = \arg\max_{\theta} \ \texttt{modified\_log}(\theta, \lambda^{t+1}) \qquad \rightarrow \text{Maximization Step}$$

Observe the dataset

Observe the dataset

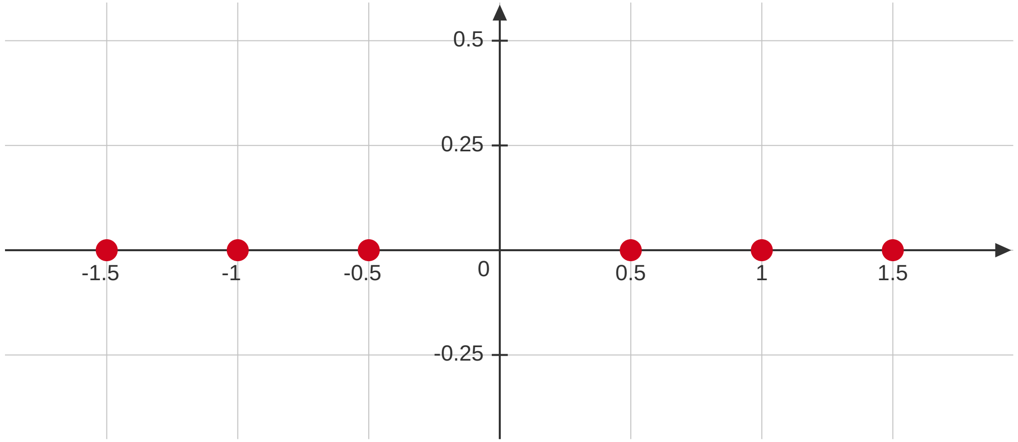
Consider the following one-dimensional dataset:

$$X = \begin{bmatrix} -1.5 & -1 & -0.5 & 0.5 & 1 & 1.5 \end{bmatrix}$$

Observe the dataset

Consider the following one-dimensional dataset:

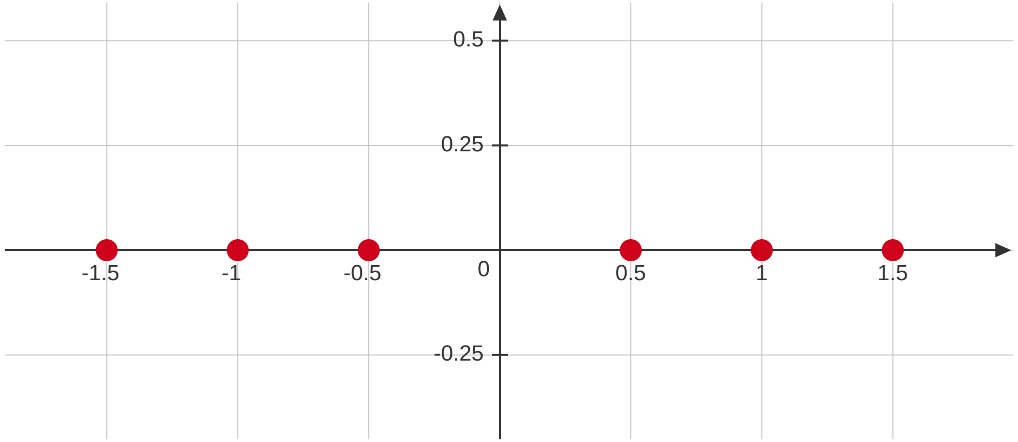
$$X = \begin{bmatrix} -1.5 & -1 & -0.5 & 0.5 & 1 & 1.5 \end{bmatrix}$$



Observe the dataset

Consider the following one-dimensional dataset:

$$X = \begin{bmatrix} -1.5 & -1 & -0.5 & 0.5 & 1 & 1.5 \end{bmatrix}$$



We can initialize θ^0 as,

$$\begin{aligned} \mu_1^0 &= -0.667 \\ \mu_2^0 &= 0.667 \\ {\sigma^2}_1^0 &= 0.722 \\ {\sigma^2}_2^0 &= 0.722 \\ \pi_1^0 &= 0.5 \\ \pi_2^0 &= 0.5 \end{aligned}$$

E-Step: Find the values - λ_k^i

E-Step: Find the values - λ_k^i

$$\widehat{\lambda}_k^{i\ MML} = \frac{\left(\frac{1}{\sqrt{2\pi}\sigma_k} e^{\frac{-(x_i-\mu_k)^2}{2\sigma_k^2}}\right) * \pi_k}{\sum_{k=1}^{K}\left(\frac{1}{\sqrt{2\pi}\sigma_k} e^{\frac{-(x_i-\mu_k)^2}{2\sigma_k^2}} * \pi_k\right)}$$

E-Step: Find the values - λ_k^i

$$\widehat{\lambda}_k^{i\ MML} = \frac{\left(\frac{1}{\sqrt{2\pi}\sigma_k} e^{\frac{-(x_i-\mu_k)^2}{2\sigma_k^2}}\right) * \pi_k}{\sum_{k=1}^{K} \left(\frac{1}{\sqrt{2\pi}\sigma_k} e^{\frac{-(x_i-\mu_k)^2}{2\sigma_k^2}} * \pi_k\right)}$$

i

k

λ	1	2	3	4	5	6
1	λ_1^1	λ_1^2	λ_1^3	λ_1^4	λ_1^5	λ_1^6
2	λ_2^1	λ_2^2	λ_2^3	λ_2^4	λ_2^5	λ_2^6

E-Step: Find the values - λ_k^i

$$\widehat{\lambda}_k^{i\ MML} = \frac{\left(\frac{1}{\sqrt{2\pi}\sigma_k} e^{\frac{-(x_i-\mu_k)^2}{2\sigma_k^2}}\right) * \pi_k}{\sum_{k=1}^{K} \left(\frac{1}{\sqrt{2\pi}\sigma_k} e^{\frac{-(x_i-\mu_k)^2}{2\sigma_k^2}} * \pi_k\right)}$$

$$\lambda_1^1 = \frac{\left(\frac{1}{(0.849)\sqrt{2\pi}} e^{\frac{-(-1.5+0.667)^2}{2(0.722)}}\right) * 0.5}{\left(\frac{1}{(0.849)\sqrt{2\pi}} e^{\frac{-(-1.5+0.667)^2}{2(0.722)}}\right) * 0.5 + \left(\frac{1}{(0.849)\sqrt{2\pi}} e^{\frac{-(1.5-0.667)^2}{2(0.722)}}\right) * 0.5}$$

$i \longrightarrow$; $k \downarrow$

λ	1	2	3	4	5	6
1	λ_1^1	λ_1^2	λ_1^3	λ_1^4	λ_1^5	λ_1^6
2	λ_2^1	λ_2^2	λ_2^3	λ_2^4	λ_2^5	λ_2^6

E-Step: Find the values - λ_k^i

$$\widehat{\lambda}_k^{i\ MML} = \frac{\left(\frac{1}{\sqrt{2\pi}\sigma_k} e^{\frac{-(x_i-\mu_k)^2}{2\sigma_k^2}}\right) * \pi_k}{\sum_{k=1}^{K}\left(\frac{1}{\sqrt{2\pi}\sigma_k} e^{\frac{-(x_i-\mu_k)^2}{2\sigma_k^2}} * \pi_k\right)}$$

$$\lambda_1^1 = \frac{\left(\frac{1}{(0.849)\sqrt{2\pi}} e^{\frac{-(-1.5+0.667)^2}{2(0.722)}}\right) * 0.5}{\left(\frac{1}{(0.849)\sqrt{2\pi}} e^{\frac{-(-1.5+0.667)^2}{2(0.722)}}\right) * 0.5 + \left(\frac{1}{(0.849)\sqrt{2\pi}} e^{\frac{-(1.5-0.667)^2}{2(0.722)}}\right) * 0.5}$$

i

k

λ	1	2	3	4	5	6
1	λ_1^1	λ_1^2	λ_1^3	λ_1^4	λ_1^5	λ_1^6
2	λ_2^1	λ_2^2	λ_2^3	λ_2^4	λ_2^5	λ_2^6

λ	1	2	3	4	5	6
1	0.940	0.863	0.715	0.284	0.136	0.059
2	0.059	0.136	0.284	0.715	0.863	0.940

M-step: Find the parameters of θ^t

M-step: Find the parameters of θ^t

The closed form expressions of the involved parameters is given by,

$$\widehat{\mu}_k^{MML} = \frac{\sum_{i=1}^{n} \lambda_k^i x_i}{\sum_{i=1}^{n} \lambda_k^i}$$

$$\widehat{\sigma^2}_k^{MML} = \frac{\sum_{i=1}^{n} \lambda_k^i (x_i - \widehat{\mu}_k^{MML})^2}{\sum_{i=1}^{n} \lambda_k^i}$$

$$\widehat{\pi}_k^{MML} = \frac{\sum_{i=1}^{n} \lambda_k^i}{n}$$

M-step: Find the values - μ_k

M-step: Find the values - μ_k

$$\widehat{\mu}_1^{MML} = \frac{\left(\lambda_1^1 x_1 + \lambda_1^2 x_2 + \lambda_1^3 x_3 + \lambda_1^4 x_4 + \lambda_1^5 x_5 + \lambda_1^6 x_6\right)}{\left(\lambda_1^1 + \lambda_1^2 + \lambda_1^3 + \lambda_1^4 + \lambda_1^5 + \lambda_1^6\right)}$$

$$= \frac{(0.940)(-1.5) + (0.863)(-1) + (0.715)(-0.5) + (0.284)(0.5) + (0.136)(1) + (0.059)(1.5)}{0.940 + 0.863 + 0.715 + 0.284 + 0.136 + 0.059}$$

$$= -0.755$$

M-step: Find the values - μ_k

$$\widehat{\mu}_1^{MML} = \frac{\left(\lambda_1^1 x_1 + \lambda_1^2 x_2 + \lambda_1^3 x_3 + \lambda_1^4 x_4 + \lambda_1^5 x_5 + \lambda_1^6 x_6\right)}{\left(\lambda_1^1 + \lambda_1^2 + \lambda_1^3 + \lambda_1^4 + \lambda_1^5 + \lambda_1^6\right)}$$

$$= \frac{(0.940)(-1.5) + (0.863)(-1) + (0.715)(-0.5) + (0.284)(0.5) + (0.136)(1) + (0.059)(1.5)}{0.940 + 0.863 + 0.715 + 0.284 + 0.136 + 0.059}$$

$$= -0.755$$

$$\widehat{\mu}_2^{MML} = \frac{\left(\lambda_2^1 x_1 + \lambda_2^2 x_2 + \lambda_2^3 x_3 + \lambda_2^4 x_4 + \lambda_2^5 x_5 + \lambda_2^6 x_6\right)}{\left(\lambda_2^1 + \lambda_2^2 + \lambda_2^3 + \lambda_2^4 + \lambda_2^5 + \lambda_2^6\right)}$$

$$= \frac{(0.059)(-1.5) + (0.136)(-1) + (0.284)(-0.5) + (0.715)(0.5) + (0.863)(1) + (0.940)(1.5)}{0.059 + 0.136 + 0.284 + 0.715 + 0.863 + 0.940}$$

$$= 0.755$$

M-step: Find the values - σ_k^2

M-step: Find the values - σ_k^2

$$\widehat{\sigma_1^2}^{MML} = \frac{\sum_{i=1}^{n} \lambda_1^i (x_i - \widehat{\mu}_1^{MML})^2}{\sum_{i=1}^{n} \lambda_1^i}$$

$$= \frac{0.940(-1.5+0.755)^2 \; + \; 0.863(-1+0.755)^2 \; + \; 0.715(-0.5+0.755)^2 + 0.284(0.5+0.755)^2 + 0.136(1+0.755)^2 + 0.059(1.5+0.755)^2}{0.940+0.863+0.715+0.284+0.136+0.059}$$

$$= 0.596$$

M-step: Find the values - σ_k^2

$$\begin{aligned}
\widehat{\sigma_1^2}^{MML} &= \frac{\displaystyle\sum_{i=1}^{n} \lambda_1^i (x_i - \widehat{\mu}_1^{MML})^2}{\displaystyle\sum_{i=1}^{n} \lambda_1^i} \\
&= \frac{0.940(-1.5+0.755)^2 \; + \; 0.863(-1+0.755)^2 \; + \; 0.715(-0.5+0.755)^2 + 0.284(0.5+0.755)^2 + 0.136(1+0.755)^2 + 0.059(1.5+0.755)^2}{0.940+0.863+0.715+0.284+0.136+0.059} \\
&= 0.596
\end{aligned}$$

$$\begin{aligned}
\widehat{\sigma_2^2}^{MML} &= \frac{\displaystyle\sum_{i=1}^{n} \lambda_2^i (x_i - \widehat{\mu}_2^{MML})^2}{\displaystyle\sum_{i=1}^{n} \lambda_2^i} \\
&= \frac{0.059(-1.5-0.755)^2 \; + \; 0.136(-1-0.755)^2 \; + \; 0.284(-0.5-0.755)^2 + 0.715(0.5-0.755)^2 + 0.863(1-0.755)^2 + 0.940(1.5-0.755)^2}{0.059+0.136+0.284+0.715+0.863+0.940} \\
&= 0.596
\end{aligned}$$

M-step: Find the values - π_k

M-step: Find the values - π_k

$$\widehat{\pi}_1^{MML} = \frac{\sum_{i=1}^{n} \lambda_1^i}{n} = \frac{0.940 + 0.863 + 0.715 + 0.284 + 0.0.136 + 0.059}{6} = 0.5$$

$$\widehat{\pi}_2^{MML} = \frac{\sum_{i=1}^{n} \lambda_2^i}{n} = \frac{0.059 + 0.136 + 0.284 + 0.715 + 0.863 + 0.940}{6} = 0.5$$

M-step: Find the values - π_k

$$\widehat{\pi}_1^{MML} = \frac{\sum_{i=1}^{n} \lambda_1^i}{n} = \frac{0.940 + 0.863 + 0.715 + 0.284 + 0.0.136 + 0.059}{6} = 0.5$$

$$\widehat{\pi}_2^{MML} = \frac{\sum_{i=1}^{n} \lambda_2^i}{n} = \frac{0.059 + 0.136 + 0.284 + 0.715 + 0.863 + 0.940}{6} = 0.5$$

We will repeat the above two steps until the given convergence criteria is satisfied $(||\theta^{t+1} - \theta^t|| \leq \epsilon)$

M-step: Find the values - π_k

$$\widehat{\pi}_1^{MML} = \frac{\sum_{i=1}^{n} \lambda_1^i}{n} = \frac{0.940 + 0.863 + 0.715 + 0.284 + 0.0.136 + 0.059}{6} = 0.5$$

$$\widehat{\pi}_2^{MML} = \frac{\sum_{i=1}^{n} \lambda_2^i}{n} = \frac{0.059 + 0.136 + 0.284 + 0.715 + 0.863 + 0.940}{6} = 0.5$$

We will repeat the above two steps until the given convergence criteria is satisfied $(||\theta^{t+1} - \theta^t|| \leq \epsilon)$

For the given example we shall perform these steps for 5 iterations.After 5 iterations the change in values in negligible.

In the iteration t = 1,

E step:

λ	1	2	3	4	5	6
1	0.940	0.863	0.715	0.284	0.136	0.059
2	0.059	0.136	0.284	0.715	0.863	0.940

M step:

t	μ_1^t	μ_2^t	σ_1^t	σ_2^t	π_1^t	π_2^t
1	-0.755	0.755	0.596	0.596	0.5	0.5

In the iteration t = 2,

E step:

λ	1	2	3	4	5	6
1	0.978	0.926	0.780	0.219	0.073	0.021
2	0.021	0.073	0.219	0.780	0.926	0.978

M step:

t	μ_1^t	μ_2^t	σ_1^t	σ_2^t	π_1^t	π_2^t
2	-0.855	0.855	0.434	0.434	0.5	0.5

In the iteration t = 3,

E step:

λ	1	2	3	4	5	6
1	0.997	0.980	0.877	0.122	0.019	0.002
2	0.002	0.019	0.122	0.877	0.980	0.997

M step:

t	μ_1^t	μ_2^t	σ_1^t	σ_2^t	π_1^t	π_2^t
3	-0.943	0.943	0.275	0.275	0.5	0.5

In the iteration t = 4,

E step:

λ	1	2	3	4	5	6
1	0.999	0.998	0.968	0.031	0.001	~0
2	~0	0.001	0.031	0.968	0.998	0.999

M step:

t	μ_1^t	μ_2^t	σ_1^t	σ_2^t	π_1^t	π_2^t
4	-0.988	0.988	0.189	0.189	0.5	0.5

In the iteration t = 5,

E step:

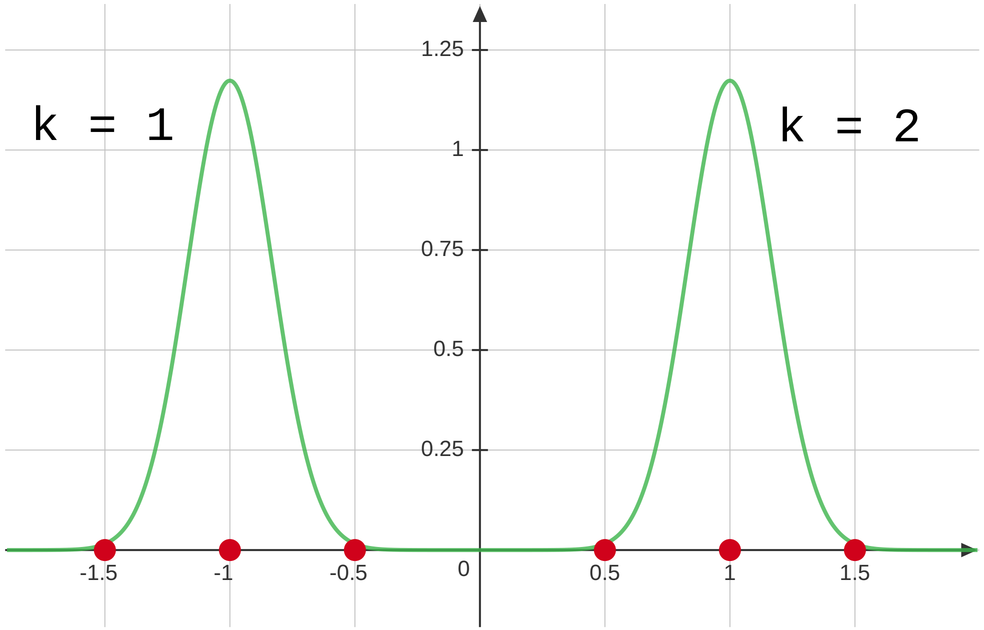
λ	1	2	3	4	5	6
1	~1	~1	~1	~0	~0	~0
2	~0	~0	~0	~1	~1	~1

M step:

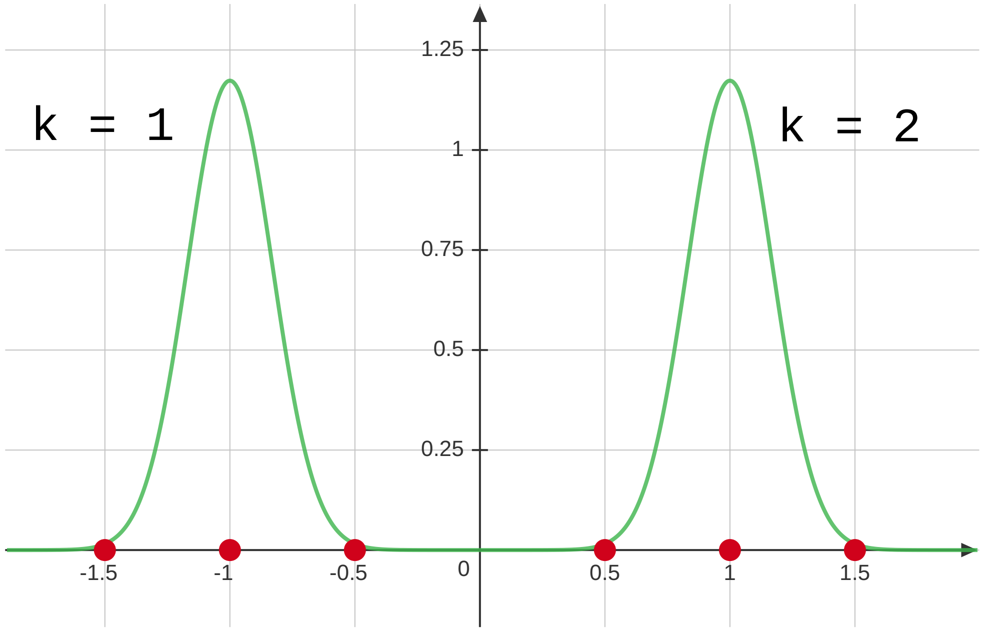
t	μ_1^t	μ_2^t	σ_1^t	σ_2^t	π_1^t	π_2^t
5	-0.998	0.998	0.170	0.170	0.5	0.5

Visual representation after convergence

Visual representation after convergence



Visual representation after convergence



The final values of θ^t are,

$$\mu_1^t = \sim -1$$
$$\mu_2^t = \sim 1$$
$${\sigma^2}_1^t = 0.17$$
$${\sigma^2}_2^t = 0.17$$
$$\pi_1^t = 0.5$$
$$\pi_2^t = 0.5$$

Q: Consider the following prior for the parameter p of the Bernoulli distribution:

$$p \sim \texttt{Beta}(3, 2)$$

The dataset is as follows:

$$\{1, 0, 1, 0, 1, 1, 1, 0, 0, 0, 1, 1, 1\}$$

Q: Consider the following prior for the parameter p of the Bernoulli distribution:

$$p \sim \texttt{Beta}(3, 2)$$

The dataset is as follows:

$$\{1, 0, 1, 0, 1, 1, 1, 0, 0, 0, 1, 1, 1\}$$

What is $\widehat{p}$, a point estimate for the parameter of the Bernoulli distribution if we use the expectation of the posterior as the method of estimation?

Q: Consider the following prior for the parameter p of the Bernoulli distribution:

$$p \sim \text{Beta}(3, 2)$$

The dataset is as follows:

$$\{1, 0, 1, 0, 1, 1, 1, 0, 0, 0, 1, 1, 1\}$$

What is $\widehat{p}$, a point estimate for the parameter of the Bernoulli distribution if we use the expectation of the posterior as the method of estimation?

A: The Beta distribution is a conjugate prior for the Bernoulli distribution.

$$\text{Beta Prior}(\alpha, \beta) \xrightarrow[\text{Bernoulli}]{\text{Data}} \text{Beta Posterior}(\alpha + n_1, \beta + n_0)$$

Q: Consider the following prior for the parameter p of the Bernoulli distribution:

$$p \sim \text{Beta}(3, 2)$$

The dataset is as follows:

$$\{1, 0, 1, 0, 1, 1, 1, 0, 0, 0, 1, 1, 1\}$$

What is $\widehat{p}$, a point estimate for the parameter of the Bernoulli distribution if we use the expectation of the posterior as the method of estimation?

A: The Beta distribution is a conjugate prior for the Bernoulli distribution.

$$\text{Beta Prior}(\alpha, \beta) \xrightarrow[\text{Bernoulli}]{\text{Data}} \text{Beta Posterior}(\alpha + n_1, \beta + n_0)$$

$$\begin{aligned} \text{Posterior} &= \text{Beta}(n_1 + \alpha, n_0 + \beta) \\ &= \text{Beta}(11, 7) \end{aligned}$$

Q: Consider the following prior for the parameter p of the Bernoulli distribution:

$$p \sim \text{Beta}(3, 2)$$

The dataset is as follows:

$$\{1, 0, 1, 0, 1, 1, 1, 0, 0, 0, 1, 1, 1\}$$

What is $\widehat{p}$, a point estimate for the parameter of the Bernoulli distribution if we use the expectation of the posterior as the method of estimation?

A: The Beta distribution is a conjugate prior for the Bernoulli distribution.

$$\text{Beta Prior}(\alpha, \beta) \xrightarrow[\text{Bernoulli}]{\text{Data}} \text{Beta Posterior}(\alpha + n_1, \beta + n_0)$$

$$\begin{aligned} \text{Posterior} &= \text{Beta}(n_1 + \alpha, n_0 + \beta) \\ &= \text{Beta}(11, 7) \end{aligned}$$

$$\begin{aligned} \text{E}[\text{Posterior}] &= \text{E}\Big[\text{Beta}(\alpha + n_1, \beta + n_0)\Big] \\ &= \frac{11}{18} \\ &= 0.61 \end{aligned}$$