Course Number: EEL 5840

UF ID: 0699 6263
Name: Fang Zhu

Homework 4:

Q1.

Solution:

$P(X|\mu) = \prod_{n=1}^{N} N(x_n|\mu,\sigma^2) \;=\; \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{\frac{-1}{2\sigma^2}(x_n-\mu)^2\}$

$\ln\big(P(X|\mu)\big) = \ln(\prod_{n=1}^{N} N(x_n|\mu,\sigma^2)) = N ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) + \sum_{n=1}^{N}(\frac{-1}{2\sigma^2}(x_n-\mu)^2)$

$\frac{\partial}{\partial\mu}\Big(\ln\big(P(X|\mu)\big)\Big) = \frac{\partial}{\partial\mu}(\ln(\prod_{n=1}^{N} N(x_n|\mu,\sigma^2))) = 0 + \frac{\partial}{\partial\mu}(\frac{-1}{2\sigma^2}\sum_{n=1}^{N}({x_n}^2 - 2\mu x_n + \mu^2)) = \frac{-1}{2\sigma^2}\sum_{n=1}^{N}(-2x_n + 2\mu) = \sum_{n=1}^{N} x_n - N\mu = 0$

So:

$\mu_{ML} = \frac{\sum_{n=1}^{N} x_n}{N}$ (1)

$P(X_N|\mu_N) = P(X|\mu)P(\mu|\mu_0) = \prod_{n=1}^{N} N(x_n|\mu,\sigma^2)\, N(\mu|\mu_0,{\sigma_0}^2)$

$\ln\big(P(X_N|\mu_N)\big) = \ln(\prod_{n=1}^{N} N(x_n|\mu,\sigma^2)) + \ln\big(N(\mu|\mu_0,{\sigma_0}^2)\big)$

$\ln\big(P(X_N|\mu_N)\big) = N ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) + \sum_{n=1}^{N}(\frac{-1}{2\sigma^2}(x_n-\mu)^2) + ln\left(\frac{1}{\sqrt{2\pi{\sigma_0}^2}}\right) + \frac{-1}{2{\sigma_0}^2}(\mu-\mu_0)^2$

$\frac{\partial}{\partial\mu}(\ln(P(X_N|\mu_N))) = \frac{\partial}{\partial\mu}\Big(\ln(\prod_{n=1}^{N} N(x_n|\mu,\sigma^2)) + \ln\big(N(\mu|\mu_0,{\sigma_0}^2)\big)\Big) = 0 + \frac{\partial}{\partial\mu}(\frac{-1}{2\sigma^2}\sum_{n=1}^{N}({x_n}^2 - 2\mu x_n + \mu^2)) + 0 + \frac{\partial}{\partial\mu}(\frac{-1}{2{\sigma_0}^2}(\mu^2 - 2\mu\mu_0 + {\mu_0}^2)) = \frac{-1}{2\sigma^2}\sum_{n=1}^{N}(-2x_n + 2\mu) + \frac{-1}{2\sigma^2}(-2\mu + 2\mu_0) = 0$ (2)

$\frac{\partial}{\partial\mu}(\ln(P(X_N|\mu_N))) = \frac{\partial}{\partial\mu}\big(N(\mu|\mu_N,{\sigma_N}^2)\big) = 0 + \frac{\partial}{\partial\mu}\left(\frac{-1}{2{\sigma_N}^2}(\mu^2 - 2\mu\mu_N + {\mu_N}^2)\right)$

$$\frac{\partial}{\partial\mu}\left(\frac{-1}{2\sigma_N{}^2}(\mu^2-2\mu\mu_N+\mu_N{}^2)\right)= \frac{-1}{2\sigma_N{}^2}(-2\mu+2\mu_N)=0$$

So: $\mu = \mu_N$ Then substitute the μ into μ_N in the function (2)

$$\frac{-1}{2\sigma^2}\sum_{n=1}^{N}(-2x_n+2\mu_N) + \frac{-1}{2\sigma_0{}^2}(2\mu_N-2\mu_0) = 0$$

$$\frac{-1}{\sigma^2}N\mu_N + \frac{1}{\sigma^2}\sum_{n=1}^{N}x_n + \frac{-1}{\sigma_0{}^2}\mu_N + \frac{1}{\sigma_0{}^2}\mu_0 = 0$$

So: $\mu_N = \frac{\sigma_0{}^2N}{\sigma_0{}^2N+\sigma^2} * \frac{\sum_{n=1}^{N}x_n}{N} + \frac{\sigma^2}{\sigma_0{}^2N+\sigma^2}\mu_0$ then substitute μ_{ML} in function (2)

So: $\mu_N = \frac{\sigma_0{}^2N}{\sigma_0{}^2N+\sigma^2}\mu_{ML} + \frac{\sigma^2}{\sigma_0{}^2N+\sigma^2}\mu_0$

According to the function above (2) we can get:

$$\ln(P(X_N|\mu_N)) = N ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)+\sum_{n=1}^{N}\left(\frac{-1}{2\sigma^2}(x_n-\mu)^2\right) + ln\left(\frac{1}{\sqrt{2\pi\sigma_0{}^2}}\right) + \frac{-1}{2\sigma_0{}^2}(\mu-\mu_0)^2$$

Because (P(X|Y)P(Y))/P(X) = P(Y|X)

Then P(X|Y)P(Y) $\propto$ P(Y|X)

Ln(P(X|Y))+LN(P(Y)) – Ln(P(X)) – Ln(P(Y|X)) = k

k is scalar.

This means that the derivative on the left equal to zero.

$$ln\left(\frac{1}{\sqrt{2\pi\sigma_N{}^2}}\right)+ \frac{-1}{2\sigma_N{}^2}(\mu- \mu_N)^2 - N ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)+\sum_{n=1}^{N}\left(\frac{-1}{2\sigma^2}(x_n-\mu)^2\right) + ln\left(\frac{1}{\sqrt{2\pi\sigma_0{}^2}}\right) + \frac{-1}{2\sigma_0{}^2}(\mu-\mu_0)^2 = k$$

$$\frac{\partial}{\partial\mu}\left(ln\left(\frac{1}{\sqrt{2\pi\sigma_N{}^2}}\right)+ \frac{-1}{2\sigma_N{}^2}(\mu- \mu_N)^2 - N ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)+\sum_{n=1}^{N}\left(\frac{-1}{2\sigma^2}(x_n-\mu)^2\right) + ln\left(\frac{1}{\sqrt{2\pi\sigma_0{}^2}}\right) + \frac{-1}{2\sigma_0{}^2}(\mu-\mu_0)^2\right) = -\frac{1}{\sigma^2}\sum_{n=1}^{N}x_n + \frac{N}{\sigma^2}\mu + \frac{1}{\sigma_0{}^2}\mu - \frac{1}{\sigma_0{}^2}\mu_0 - \frac{1}{\sigma_N{}^2}\mu + \frac{1}{\sigma_N{}^2}\mu_N = 0$$

Then the coefficients for μ should equal to zero.

So: $\frac{1}{{\sigma_N}^2} = \frac{N}{\sigma^2} + \frac{1}{{\sigma_0}^2}$

Q2:

Solution: x_n, μ, μ_0, μ_N are vectors, while σ, σ_0, σ_N are fixed number

$P(X|\mu) = \prod_{n=1}^{N} N(x_n|\mu, \Sigma) = \prod_{n=1}^{N} \frac{1}{(2\pi)^{D/2}} \frac{1}{(\Sigma)^{1/2}} \exp\{\frac{-1}{2}(x_n - \mu)^T \Sigma^{-1}(x_n - \mu)\}$

$P(\mu|\mu_0) = N(\mu|\mu_0, \Sigma_0)$

$$= \frac{1}{(2\pi)^{D/2}} \frac{1}{(\Sigma_0)^{1/2}} \exp\{\frac{-1}{2}(\mu_0 - \mu)^T {\Sigma_0}^{-1}(\mu_0 - \mu)\}$$

$$\ln\big(P(X|\mu)\big) = \ln\left(\prod_{n=1}^{N} N(x_n|\mu, \Sigma_N)\right)$$

$$= N ln\left(\frac{1}{(2\pi)^{D/2}} \frac{1}{(\Sigma)^{1/2}}\right) + \sum_{n=1}^{N}(\frac{-1}{2}(x_n - \mu)^T \Sigma^{-1}(x_n - \mu))$$

$\frac{\partial}{\partial \mu}\big(\ln\big(P(X|\mu)\big)\big) = \frac{\partial}{\partial \mu}(\ln(\prod_{n=1}^{N} N(x_n|\mu, \Sigma))) = 0 + \frac{\partial}{\partial \mu}(\frac{-1}{2}\sum_{n=1}^{N}({x_n}^T \Sigma^{-1} x_n - \mu \Sigma^{-1} {x_n}^T - \mu^T \Sigma^{-1} x_n + \mu^T \Sigma^{-1} \mu)) = \frac{-1}{2\sigma^2}\sum_{n=1}^{N} \Sigma^{-1}(-2x_n + 2\mu) = \sum_{n=1}^{N} x_n - N\mu = 0$

So:

$$\mu_{ML} = \frac{\sum_{n=1}^{N} x_n}{N} \tag{1}$$

$P(X_N|\mu_N) = P(X|\mu)P(\mu|\mu_0) = \prod_{n=1}^{N} N(x_n|\mu, \Sigma)\, N(\mu|\mu_0, \Sigma_0)$

$\ln\big(P(X_N|\mu_N)\big) = \ln(\prod_{n=1}^{N} N(x_n|\mu, \Sigma)) + \ln\big(N(\mu|\mu_0, \Sigma_0)\big)$

$\ln\big(P(X_N|\mu_N)\big) = N ln\left(\frac{1}{(2\pi)^{D/2}} \frac{1}{(\Sigma)^{1/2}}\right) + \sum_{n=1}^{N}(\frac{-1}{2}(x_n - \mu)^T \Sigma^{-1}(x_n - \mu)) + ln\left(\frac{1}{(2\pi)^{D/2}} \frac{1}{(\Sigma_0)^{1/2}}\right) + \frac{-1}{2}(\mu_0 - \mu)^T {\Sigma_0}^{-1}(\mu_0 - \mu)$

$\frac{\partial}{\partial \mu}(\ln(P(X_N|\mu_N))) = \frac{\partial}{\partial \mu}\big(\ln(\prod_{n=1}^{N} N(x_n|\mu, \Sigma)) +$

$\ln(N(\mu|\mu_0, \Sigma_0))) = \frac{-1}{2}\sum_{n=1}^{N} \Sigma^{-1}(-2x_n + 2\mu) + \frac{-1}{2}\Sigma_0{}^{-1}(2\mu - 2\mu_0) = 0$ (2)

$\frac{\partial}{\partial\mu}(\ln(P(X_N|\mu_N))) = \frac{\partial}{\partial\mu}\left(N(\mu|\mu_N, \sigma_N{}^2)\right) = 0 +$

$\frac{\partial}{\partial\mu}\left(\frac{-1}{2\sigma_N{}^2}\left(\mu_N{}^T\Sigma_N{}^{-1}\mu_N - \mu\Sigma_N{}^{-1}\mu_N{}^T - \mu^T\Sigma_N{}^{-1}\mu_N + \mu^T\Sigma_N{}^{-1}\mu\right)\right)$

$\frac{\partial}{\partial\mu}\left(\frac{-1}{2}\left(\mu_N{}^T\Sigma_N{}^{-1}\mu_N - \mu\Sigma_N{}^{-1}\mu_N{}^T - \mu^T\Sigma_N{}^{-1}\mu_N + \mu^T\Sigma_N{}^{-1}\mu\right)\right) = \frac{-1}{2}\Sigma_N{}^{-1}(2\mu - 2\mu_N) = 0$

So: $\mu = \mu_N$ Then substitute the μ into μ_N in the function (3)

$\frac{-1}{2}\Sigma^{-1}\sum_{n=1}^{N}(-2x_n + 2\mu_N) + \frac{-1}{2}\Sigma_0{}^{-1}(2\mu_N - 2\mu_0) = 0$

$-\Sigma^{-1}N\mu_N + \Sigma^{-1}\sum_{n=1}^{N}x_n + -\Sigma_0{}^{-1}\mu_N + \Sigma_0{}^{-1}\mu_0 = 0$

So: $\mu_N = \frac{\Sigma_0 N}{N\Sigma_0+\Sigma} * \frac{\sum_{n=1}^{N}x_n}{N} + \frac{\Sigma}{N\Sigma_0+\Sigma}\mu_0$ then substitute μ_{ML} in function (2)

According to the function above (2) we can get:

$\mathrm{Ln}(P(X_N|\mu_N)) = N ln\left(\frac{1}{(2\pi)^{\frac{D}{2}}}\frac{1}{(\Sigma)^{\frac{1}{2}}}\right) + \sum_{n=1}^{N}\left(\frac{-1}{2}(x_n - \mu)^T\Sigma^{-1}(x_n - \mu)\right) + ln\left(\frac{1}{(2\pi)^{\frac{D}{2}}}\frac{1}{(\Sigma_0)^{\frac{1}{2}}}\right) + \frac{-1}{2}(\mu_0 - \mu)^T\Sigma_0{}^{-1}(\mu_0 - \mu)$

Because (P(X|Y)P(Y))/P(X) = P(Y|X)

Then P(X|Y)P(Y) $\propto$ P(Y|X)

Ln(P(X|Y))+LN(P(Y)) – Ln(P(X)) – Ln(P(Y|X)) = k

k is scalar.

This means that the derivative on the left equal to zero.

$\ln(P(X_N|\mu_N)) - N ln\left(\frac{1}{(2\pi)^{\frac{D}{2}}}\frac{1}{(\Sigma)^{\frac{1}{2}}}\right) + \sum_{n=1}^{N}\left(\frac{-1}{2}(x_n - \mu)^T\Sigma^{-1}(x_n -\right.$

$\mu)\Big) + \ ln\left(\frac{1}{(2\pi)^{\frac{D}{2}}}\frac{1}{(\Sigma_0)^{\frac{1}{2}}}\right) + \frac{-1}{2}(\mu_0 - \mu)^T\Sigma_0{}^{-1}(\mu_0 - \mu)$ = k

$$\frac{\partial}{\partial\mu}\Bigg(\ln(P(X_N|\mu_N)) - \ Nln\left(\frac{1}{(2\pi)^{\frac{D}{2}}}\frac{1}{(\Sigma)^{\frac{1}{2}}}\right) + \sum_{n=1}^{N}\Big(\frac{-1}{2}(x_n -$$

$$\mu)^T\Sigma^{-1}(x_n - \mu)\Big) + \ ln\left(\frac{1}{(2\pi)^{\frac{D}{2}}}\frac{1}{(\Sigma_0)^{\frac{1}{2}}}\right) + \frac{-1}{2}(\mu_0 - \mu)^T\Sigma_0{}^{-1}(\mu_0 - \mu)\Bigg) \quad =$$

$-\Sigma^{-1}\sum_{n=1}^{N}x_n + \ N\Sigma^{-1}\mu + \ \Sigma_0{}^{-1}\mu - \Sigma_0{}^{-1}\mu_0 - \ \Sigma_N{}^{-1}\mu + \ \Sigma_N{}^{-1}\ \mu_N$ = 0

Then the coefficients for μ should equal to zero.

So: $\Sigma_N{}^{-1} \ = \ \ N\Sigma^{-1} + \ \Sigma_0{}^{-1}$

So, they are the same.

Q3:

I generate the normal distribution as likelihood with mean and standard variance of 1 and 0.1.

From the MAP solution above, we can get the function:

$$\mu_N = \frac{{\sigma_0}^2 N}{{\sigma_0}^2 N + \sigma^2}\mu_{ML} + \frac{\sigma^2}{{\sigma_0}^2 N + \sigma^2}\mu_0$$

$$\mu_N = \frac{1}{1 + \frac{\sigma^2}{N{\sigma_0}^2}}\mu_{ML} + \frac{1}{\frac{{\sigma_0}^2}{\sigma^2}N + 1}\mu_0$$

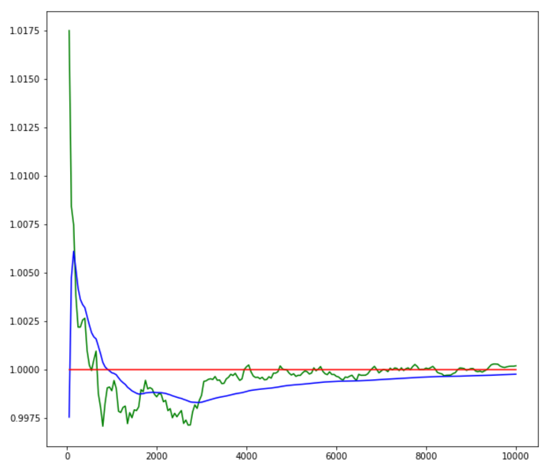
Let us suppose $\frac{\sigma^2}{{\sigma_0}^2} = K$, then:

$$\mu_N = \frac{1}{1 + \frac{K}{N}}\mu_{ML} + \frac{1}{\frac{N}{K} + 1}\mu_0$$

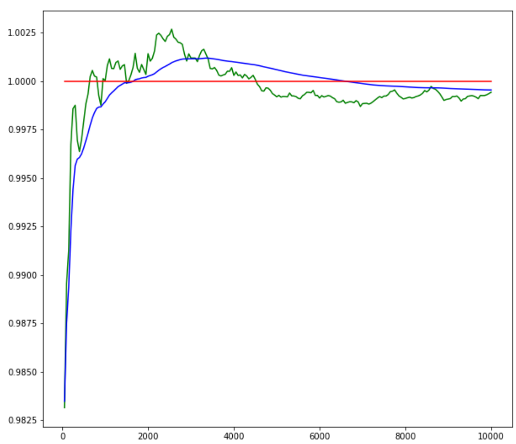
1. What happens when the prior mean is initialized to the wrong value? To the correct value?

Solution:

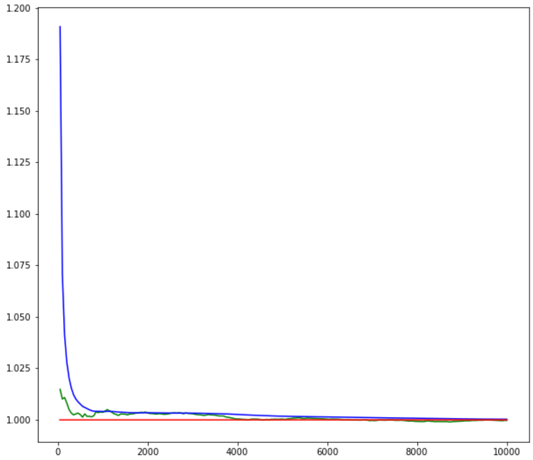
Picture of ML mean (green), MAP mean (blue) and likelihood mean (red):



Prior mean = 0



Prior mean=1



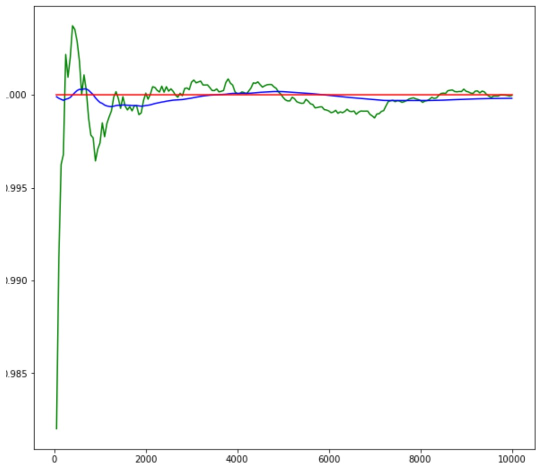
Prior = 10

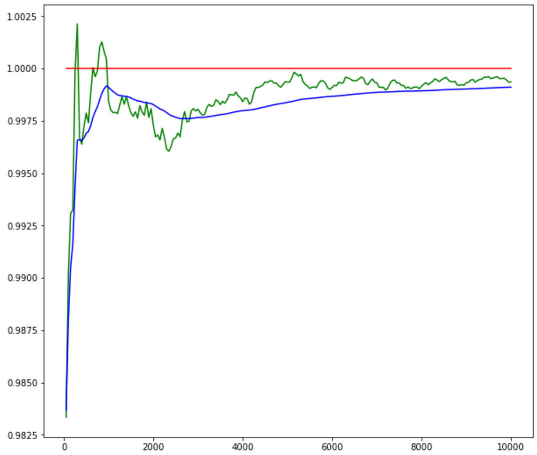
From function $\mu_N = \frac{1}{1+\frac{K}{N}}\mu_{ML} + \frac{1}{\frac{N}{K}+1}\mu_0$ we can find that to make the mean of posterior more closed to the mean of the true mean and the likelihood mean. If the mean of prior is too small, to make $\mu_N = \mu_{ML}$, N would be very large to enlarge the $\frac{1}{1+\frac{K}{N}}$. If the mean of prior is too big, to make $\mu_N = \mu_{ML}$, N would be very large to narrow the $\frac{1}{\frac{N}{K}+1}$. From the chart above, we can get the same answer. The Error between the mean of MAP and ML from 0, 1 and 10 are 0.0005, 0.0001 and 0.0006, thus to be stable closed to ML around 0.0001 for example the too big or too small prior mean requires more iteration.

2. What happens as you vary the prior variance from small to large?

Solution:

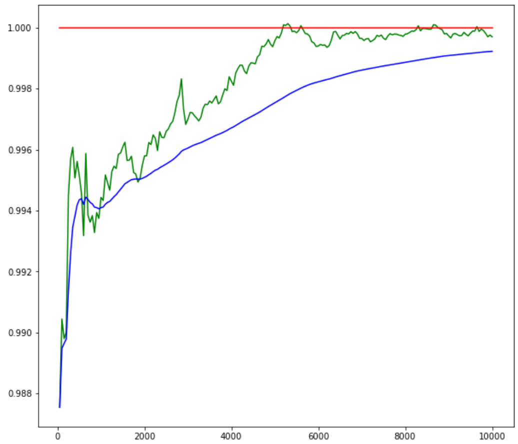
Picture of ML mean (green) and MAP mean (blue), likelihood mean (red):





Prior variance = 0.001

Prior variance=0.1



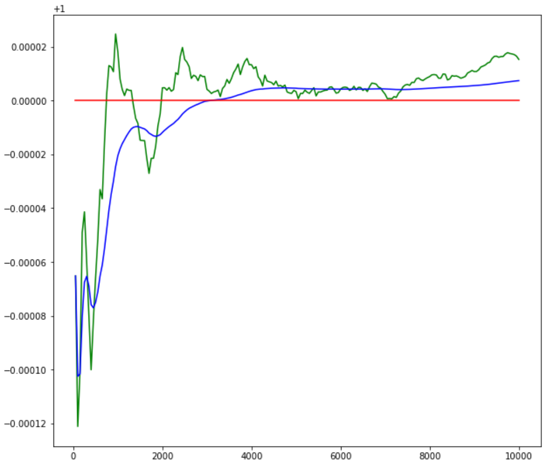
Prior variance = 100

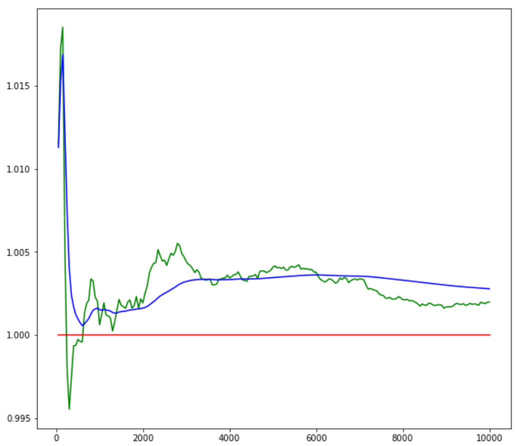
From function $\mu_N = \frac{1}{1+\frac{K}{N}}\mu_{ML} + \frac{1}{\frac{N}{K}+1}\mu_0$ we can find that to make the mean of posterior more closed to the mean of the true mean and the likelihood mean. If the variance of prior is too small (the K would be big), to make $\mu_0 = 0$ and $\mu_N = \mu_{ML}$, N would be very large to enlarge the $\frac{1}{1+\frac{K}{N}}$ and narrow the $\frac{1}{\frac{N}{K}+1}$. If the variance of prior is too big (K would be small), making $\mu_0 \to 0$ and $\mu_N \to \mu_{ML}$, N could be small. If the variance of prior is big, it means that we are not sure about this prior then we are more believed in the ML. From the chart above, we can get the same answer. The Error between the mean of MAP and ML from different prior variance 0, 1 and 10 are 0.0004, 0.0002 and 0.0001. Thus, to be stable closed to ML the too small prior variance requires more iterations.

3. What happens when the likelihood variance is varied from small to large?

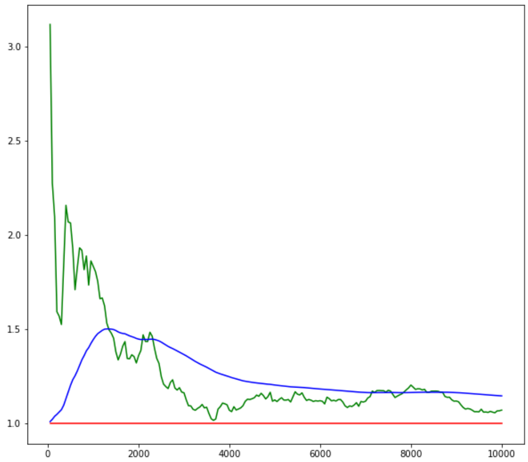
Solution:

Picture of prior mean (yellow), ML mean (green) and MAP mean (blue), likelihood mean(red):





likelihood variance=0.001 likelihood variance=0.1



likelihood variance = 10

From function $\mu_N = \frac{1}{1+\frac{K}{N}}\mu_{ML} + \frac{1}{\frac{N}{K}+1}\mu_0$ we can find that to make the mean of posterior more closed to the mean of the true mean and the likelihood mean. If the variance of prior is too small (the K would be small), making $\mu_0 \to 0$ and $\mu_N \to \mu_{ML}$, N could be small. If the variance of prior is too big (K would be big), to make $\mu_0 = 0$ and $\mu_N = \mu_{ML}$, N would be very large to enlarge the $\frac{1}{1+\frac{K}{N}}$ and narrow the $\frac{1}{\frac{N}{K}+1}$. If the variance of likelihood is big, it means that we are more believed in likelihood, then the prior. From the chart above, we can get the same answer. The Error between the mean of MAP and ML from different likelihood variance 0, 1 and 10 are 0.000008, 0.0002 and 0.007. Thus, to be stable closed to ML the too small likelihood variance requires more iterations.

4. How do the initial values of the prior mean, prior variance, and likelihood variance interact to affect the final estimate of the mean?

Solution:

The curve of the mean of MAP with the higher prior variance and lower likelihood variance and suitable (around the true mean) prior mean would be very closed to the curve of the mean of ML faster. On the contrary, it would cost more iterations. if the prior variance is too small of the likelihood variance is too big or the likelihood mean is far from the true mean.

Code:

```
import numpy as np
import matplotlib.pyplot as plt
import math

fig = plt.figure(figsize=(10,20))

mean = 1
sigma = 10
u_prior = 1
sigma_prior = 0.1
step = 50
group = 200

box = np.zeros(step*group)
U_ML = np.zeros(group)
U_MAP = np.zeros(group)
U = mean*np.ones(group)
S_ML = np.zeros(group)
S_MAP = np.zeros(group)
S = sigma*np.ones(group)
U_P = np.ones(group)*u_prior
```

```
S_P = np.ones(group)*sigma_prior

for it in range(group):
    for i in range(step):
        box[i+it*step] = np.random.normal(mean,sigma,1)
    n = (it+1)*step
    data = np.zeros(n)
    data = box[0:(it+1)*step]
    # ML
    u_ml = sum(data)/n
    I = np.ones([1,n])
    u_I = np.matrix(u_ml * I)
    sigma_ml = math.sqrt(np.asscalar((data - u_I)@(data -
u_I).T/n))
    # MAP
    u_map =
(u_prior*(sigma**2))/(n*sigma_prior**2+sigma**2)+(u_ml*n*(s
igma_prior**2))/(n*sigma_prior**2+sigma**2) # u_map
    sigma_map =
math.sqrt(1/((1/(sigma_prior**2)+n/(sigma**2))))
    print('1. The mean of the data: '+str(mean))
    print('2. The sigma of the data: '+str(sigma))
```

```
    print('3. The mean of the prior: '+str(u_prior))
    print('4. The sigma of the prior: '+str(sigma_prior))
    print('5. The mean of the ML: '+str(u_ml))
    print('6. The sigma of the ML: '+str(sigma_ml))
    print('7. The mean of the MAP: '+str(u_map))
    print('8. The sigma of the MAP: '+str(sigma_map))
    print('9. Error between the mean of MAP and ML: '+str(abs(u_map-u_ml)))
    print(' ')

print('************************************************************
*****')
    print(' ')
    # update
    u_prior = u_map
    sigma_prior = sigma_map
    U_ML[it] = u_ml
    U_MAP[it] = u_map
    S_ML[it] = sigma_ml
    S_MAP[it] = sigma_map
```

```
p1 = fig.add_subplot(*[2,1,1])
t = np.arange(step,step+step*group,step)
p1.plot(t,U_ML, 'g')
p1.plot(t,U_MAP,'b')
p1.plot(t,U,'r')
#p1.plot(t,U_P,'y')

p2 = fig.add_subplot(*[2,1,2])
#p2.plot(t,S_ML, 'g')
p2.plot(t,S_MAP,'b')
p2.plot(t,S,'r')
p2.plot(t,S_P,'y')
```