Solution to analysis in Home Assignment 1

Lizi Teng + lizi(cid)

Analysis

In this report I will present my independent analysis of the questions related to home assignment 1. I have discussed the solution with Qun Zhang, but I swear that the analysis written here are my own.

1 Properties of random variables

a)

We know that $X \sim N(\mu, \sigma^2)$, so $p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$.
i)

$$
\begin{aligned}
E[x] &= \int_{-\infty}^{\infty} xp(x)dx \\
&= \int_{-\infty}^{\infty} \frac{x}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\
assume \quad t &= \frac{x-u}{2\sigma} \\
&= \frac{2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (2\sigma t + \mu) \exp(-2t^2) dt \\
&= \frac{2}{\sqrt{2\pi}} (2\sigma \int_{-\infty}^{\infty} t \exp(-2t^2) dt + \mu \int_{-\infty}^{\infty} \exp(-2t^2) dt)
\end{aligned} \tag{1}
$$

since $texp(-2t^2)$ is a odd function,the integration is 0 from negetive to Positive infinity, and that$\int_{-\infty}^{\infty} exp(tx^2) = \frac{\sqrt{\pi}}{\sqrt{-t}}$,the result is :

$$
\begin{aligned}
E[x] &= \frac{2}{\sqrt{2\pi}} * \frac{\sqrt{2\pi}}{2} * \mu \\
&= \mu
\end{aligned} \tag{2}
$$

ii)

$$
\begin{aligned}
Var[x] = E[(x-\mu)^2] &= \int_{-\infty}^{\infty} (x-\mu)^2 p(x) dx \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x-\mu)^2 \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\
& assume \quad t = x-\mu \\
&= -\frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t \exp(-\frac{1}{2\sigma^2}t^2) d(-\frac{1}{2\sigma^2}) \\
&= -\frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t d(\exp(-\frac{1}{2\sigma^2}t^2)) \\
& using \quad integration \quad by \quad parts \\
&= -\frac{\sigma}{\sqrt{2\pi}} (\underbrace{t \exp(-\frac{1}{2\sigma^2}t^2)|_{-\infty}^{\infty}}_{0} - \int_{-\infty}^{\infty} \exp(-\frac{1}{2\sigma^2}t^2) dt) \\
&= \sigma^2
\end{aligned} \tag{3}
$$

b)

i)

$$
\begin{aligned}
E[z] &= \int_{-\infty}^{\infty} Aq * p(q) dq \\
&= A \int_{-\infty}^{\infty} q * p(q) dq \\
&= AE[q]
\end{aligned} \tag{4}
$$

ii)

$$
\begin{aligned}
cov[z] &= E[[z - E[z]][z - E[z]]^T] \\
&= E[[Aq - AE[q]][Aq - AE[q]]^T] \\
with &\quad : (AB)^T = B^T A^T \\
&= AE[[q - E[q]][q - E[q]]^T]A^T \\
&= Acov[q]A^T
\end{aligned} \tag{5}
$$

c)

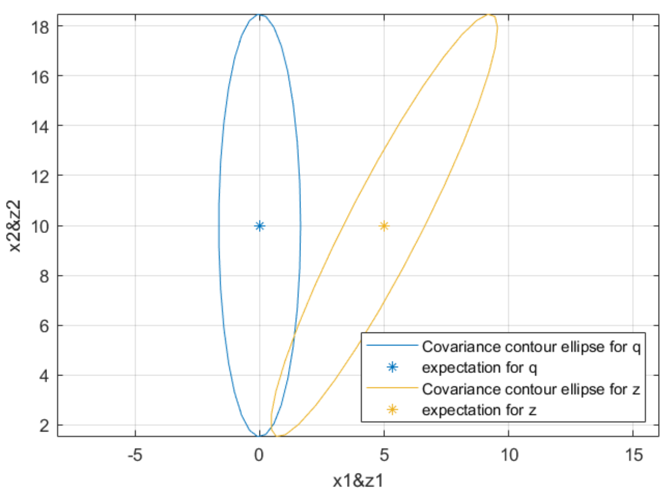

Figure 1

As we can see in figure, both the size and the position of the elipse change.The new mean is equal to $A * mean_q$, and it cause the change of the center of the elipse. We can also see the change of the elipse as an affine transition, the affine matrix A contains the rotation and scale.

The individual components started to have correlation between each other after the transformation, the distribution of one component can be different given another component, and the covariance matrix changed from:

$$\begin{bmatrix} 0.3 & 0 \\ 0 & 8 \end{bmatrix} \tag{6}$$

To:

$$\begin{bmatrix} 2.5 & 4 \\ 4 & 8 \end{bmatrix} \tag{7}$$

There are nonzero values on the off-diagonal of the A matrix so there would be nonzero values on the off-diagonal of the Cov matrix. And the components in q thus are not independent.

2 Transformation of random variables

a)

Firstly, $z = 3 * x$, it's a linear function, so we may use the function affineGaussianTransform to get the analytical result. And then we use function approxGaussianTransform to get the numerical result.Here is the result plot:

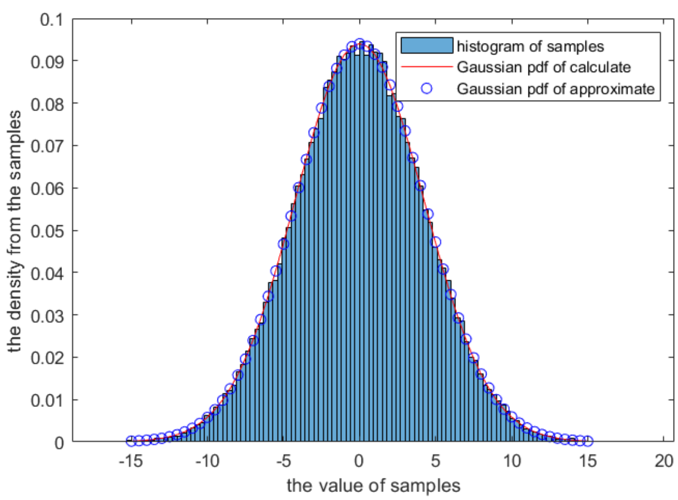

Figure 2: The numerical result and the analytical results for both calculated and approximated match well

The different approximation both fit the calculated result well, because when only linear transformation happens to Gaussian distribution, it's still a Gaussian distribution, meaning p(z) is a Gaussian distribution.
In the next figure, I tried to get 10000 samples instead of 100000. And the approximation seems worse fitted than the Figure 2.

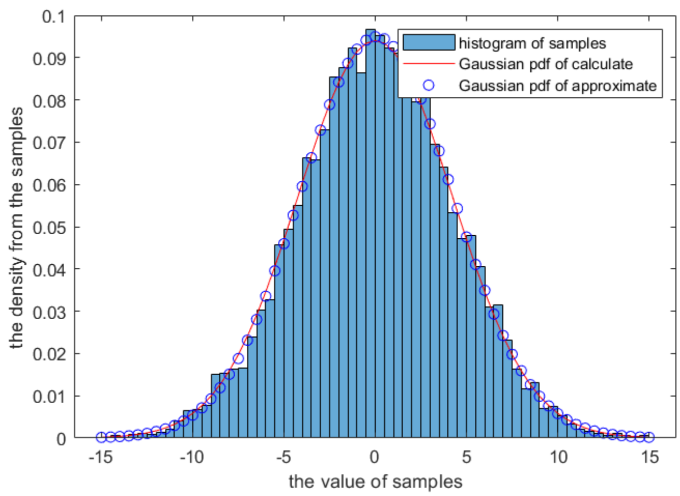

Figure 3: The numerical result and the analytical results for both calculated and approximated match well

b)

Firstly, we have to determine $p(z)$ by assuming that it is Gaussian and calculating $E[z]$ and $Var[z]$ analytically.

$$
\begin{aligned}
E[z] &= \int_{-\infty}^{\infty} x^3 p(x) dx \\
&= \int_{-\infty}^{\infty} x^3 \frac{1}{\sqrt{4\pi}} \exp(-\frac{x^2}{4}) dx \\
&= \frac{8}{\sqrt{4\pi}} \int_{-\infty}^{\infty} -(\frac{x^2}{4}) \exp(-(\frac{x^2}{4})) d(-(\frac{x^2}{4})) \\
&= \frac{8}{\sqrt{4\pi}} (-\frac{x^2}{4} - 1) exp(-\frac{x^2}{4}) |_{-\infty}^{\infty} \\
&= 0 \\
Var(z) &= E[(z - E[z])^2] \\
&= \frac{1}{2\sqrt{\pi}} \int_{-\infty}^{\infty} x^6 exp(-\frac{x^2}{4}) dx \\
& Then \quad I \quad calculate \quad with \quad integral - calculator : \\
&= (120\sqrt{\pi} erf(\frac{q}{2}) + exp(-\frac{q^2}{4}(-2q^5 - 20q^3 - 120q))) * \frac{1}{\sqrt{\pi}} |_{-\infty}^{\infty} \\
&= 120
\end{aligned} \tag{8}
$$

Then I have the plot as the same steps in 2a).And the results of numerical and analytical don't match:

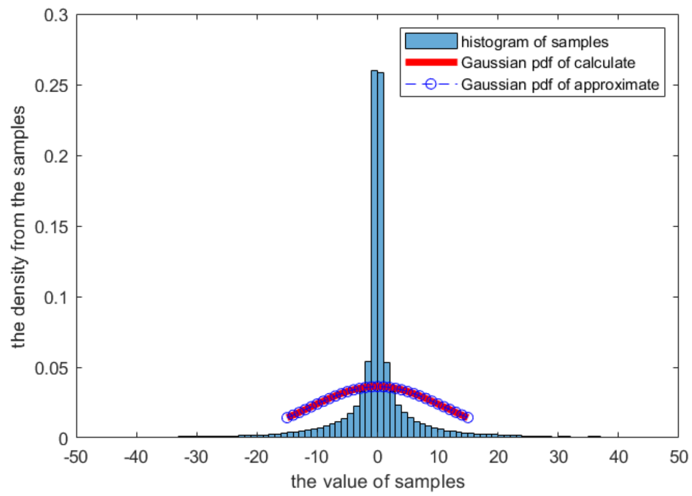

Figure 4: The numerical result and the analytical results for both calculated and approximated don't match

With a nonlinear transition:$z = x^3$, the p(z) is no longer a Gaussian distribution, so when we try to get plot from Gaussian pdfs, we have different result as the numerical one. And we have no need to compare the result of different numbers of samples, because they won't match.

c)

When $z = 3x$, it's a linear transition between x and z, and for Gaussian distribution, the p(z) is still a Gaussian distribution after transition, so we can match p(z) well with normpdf.
When $z = x^3$, it's a nonlinear transition between x and z, and p(z) is no longer a Gaussian distribution, so we can not match the plots.

3 Understanding the conditional density

a)

If we know $p(x)$ then we are able to describe $p(y)$.
Now we calculate the distribution $p(y)$:

$$\begin{aligned} p(y) &= \int_{-\infty}^{\infty} p(y, x) \\ &= \int_{-\infty}^{\infty} p(y|x)p(x)dx \end{aligned} \tag{9}$$

because $y = h(x) + r$, we get $p(y|x) \sim N(h(x), \sigma_r^2)$.

$$\begin{aligned} p(y) &= \int_{-\infty}^{\infty} N(y; h(x), \sigma_r^2)p(x)dx \\ &= \int_{-\infty}^{\infty} p(x)\frac{1}{\sqrt{2\pi}\sigma_r} \exp\left(-\frac{(y - h(x))^2}{2\sigma_r^2}\right) dx \end{aligned} \tag{10}$$

But if we don't know the distribution of x, we cannot describe p(y).

b)

As we already get from question a), We can always describe $p(y|x)$ when we know h(x) and the distribution of r.

$$\begin{aligned} p(y|x) &= N(y; h(x), \sigma_r^2) \\ &= \frac{1}{\sqrt{2\pi}\sigma_r} \exp\left(-\frac{(y - h(x))^2}{2\sigma_r^2}\right) \end{aligned} \tag{11}$$

c)

We can always get $p(y|x)$ and the result is $p(y|x) = N(y; H(x), \sigma_r^2) = \frac{1}{\sqrt{2\pi}\sigma_r} \exp\left(-\frac{(y-h(x))^2}{2\sigma_r^2}\right)$, and when $p(x)$ is determined, we can get $p(y) = \int_{-\infty}^{\infty} p(x)\frac{1}{\sqrt{2\pi}\sigma_r} \exp\left(-\frac{(y-Hx)^2}{2\sigma_r^2}\right) dx$.

d)

When we have the $x \sim N(\mu_x, \sigma_x^2)$, we can get solutions for all the previous questions.
For the nonlinear function $h(x)$, the result of p(y) and $p(y|x)$:

$$
\begin{aligned}
p(y|x) &= \frac{1}{\sqrt{2\pi}\sigma_r} \exp\left(-\frac{(y-h(x))^2}{2\sigma_r^2}\right) dx \\
p(y) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{(x-\mu_x)^2}{2\sigma_x^2}\right) \frac{1}{\sqrt{2\pi}\sigma_r} \exp\left(-\frac{(y-h(x))^2}{2\sigma_r^2}\right) dx
\end{aligned}
\tag{12}
$$

For the linear functions Hx,

$$
\begin{aligned}
p(y|x) &= \frac{1}{\sqrt{2\pi}\sigma_r} \exp\left(-\frac{(y-Hx)^2}{2\sigma_r^2}\right) dx \\
p(y) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{(x-\mu_x)^2}{2\sigma_x^2}\right) \frac{1}{\sqrt{2\pi}\sigma_r} \exp\left(-\frac{(y-Hx)^2}{2\sigma_r^2}\right) dx
\end{aligned}
\tag{13}
$$

And because function is linear, we also have the solution:

$$
\begin{aligned}
\sigma_y^2 &= H^2\sigma_x^2 + \sigma_r^2 \\
\mu_y &= H\mu_x \\
y &\sim N(y; \mu_y, \sigma_y^2)
\end{aligned}
\tag{14}
$$

e)

Firstly, I create three distributions, a uniform distribution for x, which is U(3,9); a normal distribution for x which is N(1,4);a normal distribution for which is N(0,9);
And two functions, a linear one $h(x)_l = 6 * x$, a nonlinear one $h(x)_{nl} = x^3$.

Then I calculate the analytical and numerical result for p(y) with linear h(x) and x $\sim$ uniform distribution(result 1); p(y) with nonlinear h(x) and x $\sim$ uniform distribution(result 2); p(y) with linear h(x) and x $\sim$ normal distribution(result 3); p(y) with nonlinear h(x) and x $\sim$ normal distribution(result 4). And they all fit pretty well in my graph:

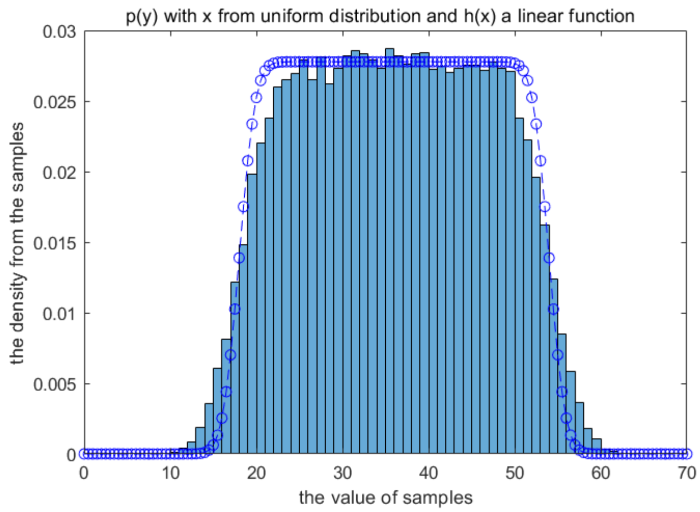

Figure 5: result 1

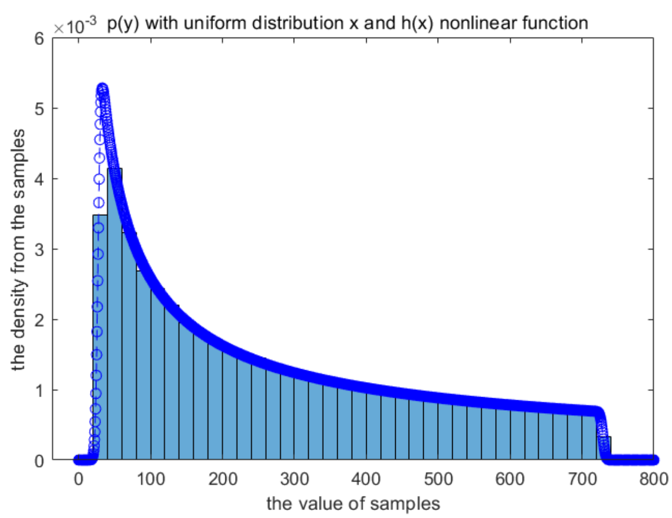

Figure 6: result 2

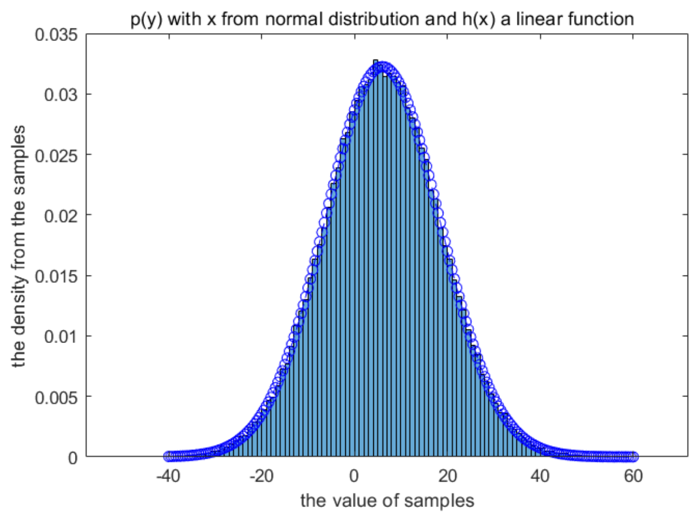

Figure 7: result 3

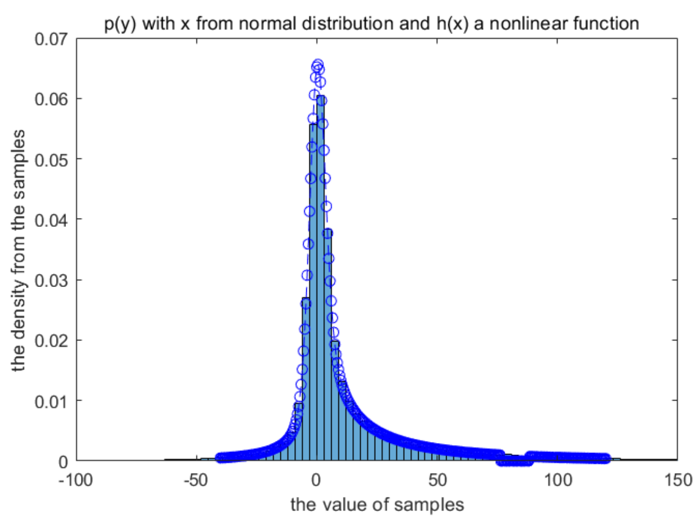

Figure 8: result 4

And I also get two samples of x from uniform distribution and normal distribution. And then get the $P(y|x)$ for both linear and nonlinear functions,they all fit pretty well in my graph:

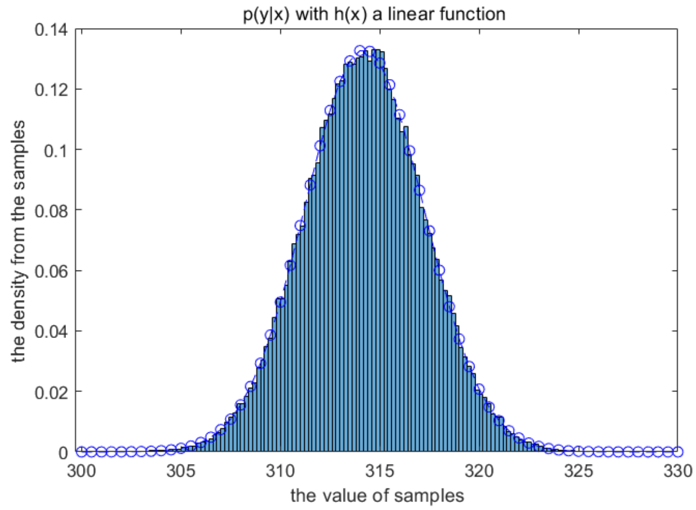

Figure 9: result 5

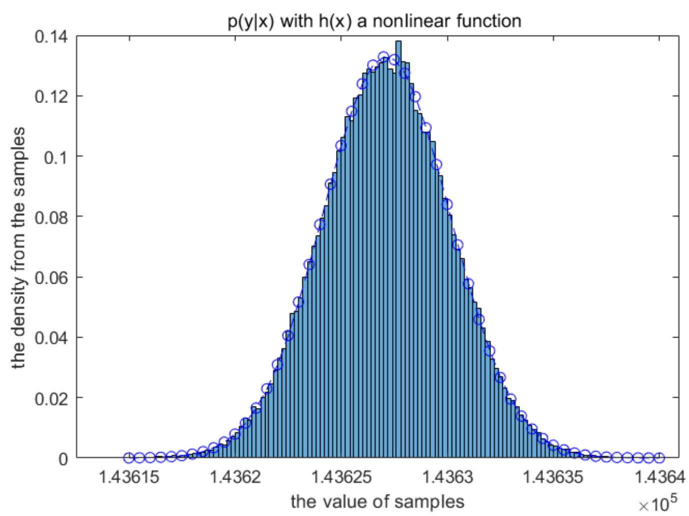

Figure 10: result 6

4 MMSE and MAP estimators

a)

The plot is as followed:

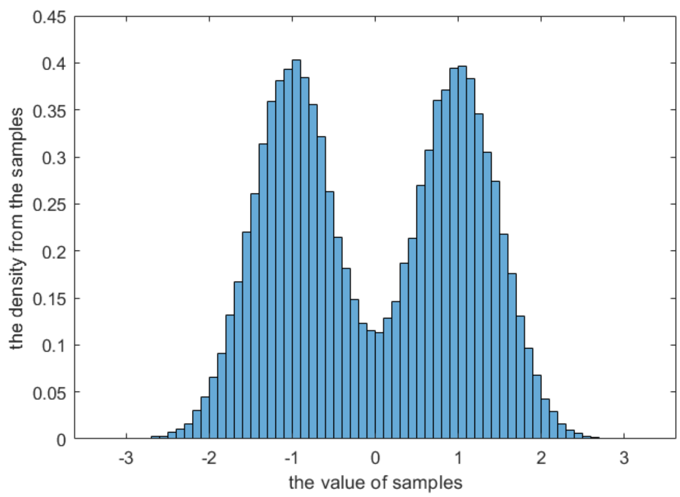

Figure 11: This figure shows that the distribution looks like two Gaussian distribution added together.

b)

I guess $\theta = 1$.
As we know from Bayesian and the problem,

$$
\begin{gathered}
p(\theta|y) = p(y|\theta)p(\theta)/p(y) \\
p(\theta = 1) = p(\theta = -1) = 0.5
\end{gathered}
\tag{15}
$$

we know that $p(\theta|y) \propto p(y|\theta)$,and we can get $p(y|\theta) = N(y; \theta, \sigma^2)$, due to the property of normal distribution, we can clearly know $N(0.7; 1, \sigma_w^2) > N(0.7; -1, \sigma_w^2)$. So we can get $p(0.7|1) > p(0.7|-1)$, so that $p(\theta = 1|0.7) > p(\theta = -1|0.7)$.And we can guess that $\theta = 1$

c)

Firstly, $p(y|\theta) = N(y;\theta,\sigma^2)$.
Because θ is a discrete variable, we have:

$$\begin{aligned}
p(y) &= \sum p(y|\theta)p(\theta) \\
&= p(y,\theta=1) + p(y,\theta=-1) \\
&= p(y|\theta=1)p(\theta=1) + p(y|\theta=-1)p(\theta=-1) \\
&= 0.5N(y;1,\sigma^2) + 0.5N(y;-1,\sigma^2) \\
&= 0.5\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-(y-1)^2}{2\sigma^2}} + 0.5\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-(y+1)^2}{2\sigma^2}}
\end{aligned} \tag{16}$$

d)

$$\begin{aligned}
p(\theta=1|y) &= \frac{p(y|\theta=1)p(\theta=1)}{p(y)} \\
&= \frac{0.5 * \frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-(y-1)^2}{2\sigma^2}}}{0.5\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-(y-1)^2}{2\sigma^2}} + 0.5\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-(y+1)^2}{2\sigma^2}}} \\
&= \frac{e^{\frac{-(y-1)^2}{2\sigma^2}}}{e^{\frac{-(y-1)^2}{2\sigma^2}} + e^{\frac{-(y+1)^2}{2\sigma^2}}} \\
&= \frac{1}{e^{-\frac{2y}{\sigma^2}} + 1} \\
p(\theta=-1|y) &= \frac{p(y|\theta=-1)p(\theta=-1)}{p(y)} \\
&= \frac{0.5 * \frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-(y+1)^2}{2\sigma^2}}}{0.5\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-(y-1)^2}{2\sigma^2}} + 0.5\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-(y+1)^2}{2\sigma^2}}} \\
&= \frac{e^{\frac{-(y+1)^2}{2\sigma^2}}}{e^{\frac{-(y-1)^2}{2\sigma^2}} + e^{\frac{-(y+1)^2}{2\sigma^2}}} \\
&= \frac{1}{e^{\frac{2y}{\sigma^2}} + 1}
\end{aligned} \tag{17}$$

e)

$$
\begin{aligned}
\hat{\theta}_{MMSE} &= 1 * p(\theta = 1|y) + (-1) * p(\theta = -1|y) \\
&= \frac{N(1,\sigma^2) - N(-1,\sigma^2)}{N(1,\sigma^2) + N(-1,\sigma^2)} \\
&= \frac{e^{\frac{2y}{\sigma^2}} - 1}{e^{\frac{2y}{\sigma^2}} + 1} \\
&= tanh(\frac{y}{\sigma^2})
\end{aligned} \tag{18}
$$

f)

In the MAP estimator, we have to choose the θ that make the posterior $p(\theta|y)$ the maximum. The value of θ have two cases, so we can simply calculate the difference of the two kind of posteriors with respect to two θ. We have the $D = p(\theta = 1|y) - p(\theta = -1|y)$, so that if $D > 0$ we choose $\theta = 1$ and when $D < 0$ we choose $\theta = -1$. And the equation D is as followed:

$$
\begin{aligned}
D &= p(\theta = 1|y) - p(\theta = -1|y) \\
&= \frac{N(1,\sigma^2) - N(-1,\sigma^2)}{N(1,\sigma^2) + N(-1,\sigma^2)} \\
&= \frac{e^{\frac{2y}{\sigma^2}} - 1}{e^{\frac{2y}{\sigma^2}} + 1} \\
&= tanh(\frac{y}{\sigma^2})
\end{aligned} \tag{19}
$$

g)

We have the same equation for both MMSE and MAP estimator, but the two estimator have different meaning, the MAP get the θ that make the posterior largest, and the MMSE get the expectation of θ with the distribution $p(\theta|y)$. As you can see in the followed figure, I get the same result for MMSE and MAP, when given $y = 0.7$, we can get $\theta = 1$, both coincides with my guess. For MMSE, as in plot, when $y = 0.7$,$\hat{\theta}_{MMSE} = 1$. And for MAP, as in plot, when $y = 0.7$,$D > 0$, so $\hat{\theta}_{MAP} = 1$.

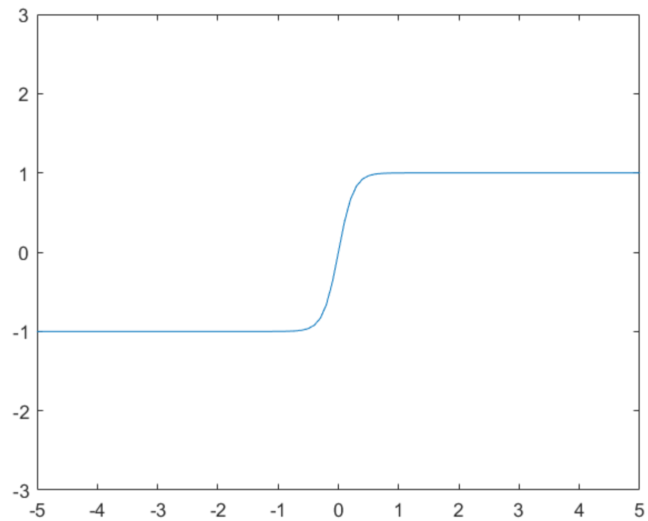

Figure 12