Solution to analysis in Home Assignment X

Yongzhao Chen (yongzhao@chalmers.se)

March 29, 2023

1 Properties of random variables

1.1 a

1.1.1 i

The definition of the expected value for a continuous random variable is:

$$\mathbb{E}[x] = \int_{-\infty}^{\infty} x f(x) dx$$

Where f(x) is the probability density function (pdf) of x. For a Gaussian random variable, the pdf is given by:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Now, we need to compute the expected value:

$$\mathbb{E}[x] = \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Let's perform the substitution:

$$t = \frac{x-\mu}{\sqrt{2}\sigma}$$

Then,

$$x = \mu + \sqrt{2}\sigma t$$
$$dx = \sqrt{2}\sigma dt$$

Now, substitute these values into the expected value integral:

$$\mathbb{E}[x] = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\mu + \sqrt{2}\sigma t) e^{-t^2} (\sqrt{2}\sigma dt) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} (\mu + \sqrt{2}\sigma t) e^{-t^2} dt$$

Now, split the integral into two parts:

$$\mathbb{E}[x] = \frac{\mu}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-t^2} dt + \frac{\sqrt{2}\sigma}{\sqrt{\pi}} \int_{-\infty}^{\infty} te^{-t^2} dt$$

Using the hint provided, we know that:

$$\int_{-\infty}^{\infty} e^{-t^2} dt = \sqrt{\pi}$$

The second integral is an odd function, and the integral over the whole real line of an odd function is zero. Therefore:

$$\int_{-\infty}^{\infty} te^{-t^2} dt = 0$$

Thus, the expected value is:$\mathbb{E}[x] = \mu + 0 = \mu$, which proves the question i).

1.1.2 ii

Using the definition of expected value:

$$Var[x] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx$$

Substitute the Gaussian pdf:

$$Var[x] = \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Use the same substitution as before:

$$t = \frac{x-\mu}{\sqrt{2}\sigma}$$

$$x = \mu + \sqrt{2}\sigma t$$
$$dx = \sqrt{2}\sigma dt$$

Substitute these values into the integral:

$$Var[x] = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Substitute the values we found earlier:

$$Var[x] = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sqrt{2}\sigma t)^2 e^{-t^2} (\sqrt{2}\sigma dt)$$

Simplify the expression:

$$Var[x] = \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} t^2 e^{-t^2} dt$$

To integrate the function $\int_{-\infty}^{\infty} t^2 e^{-t^2} dt$, let $u = t^2$, so that $du/dt = 2t$ and $dt = du/2t$. Then, we can rewrite the integral as follows:

$$\int_{-\infty}^{\infty} t^2 e^{-t^2} dt = \int_{-\infty}^{\infty} \frac{1}{2} u e^{-u} \frac{du}{\sqrt{u}}$$

Next, we can use the gamma function, $\Gamma(n) = \int_0^\infty x^{n-1} e^{-x} dx$, to simplify the integral:

$$\int_{-\infty}^{\infty} t^2 e^{-t^2} dt = \int_{0}^{\infty} \frac{1}{2} u e^{-u} \frac{du}{\sqrt{u}} + \int_{-\infty}^{0} \frac{1}{2} u e^{-u} \frac{du}{\sqrt{u}}$$

$$= \frac{1}{2} \int_0^\infty u^{1/2-1} e^{-u} du + \frac{1}{2} \int_0^\infty (-u)^{1/2-1} e^{u} du$$

Using the definition of the gamma function, we can simplify the two integrals to obtain:

$$\int_{-\infty}^{\infty} t^2 e^{-t^2} dt = \frac{1}{2}\Gamma\left(\frac{3}{2}\right) + \frac{1}{2}\Gamma\left(\frac{3}{2}\right) = \Gamma\left(\frac{3}{2}\right) = \frac{\sqrt{\pi}}{2}$$

Therefore,

$$Var[x] = \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} t^2 e^{-t^2} dt = \frac{2\sigma^2}{\sqrt{\pi}} \frac{\sqrt{\pi}}{2} = \sigma^2$$

.

Q.E.D.

1.2 b

1.2.1 1

The expected value of z, denoted by E[z], is defined as the integral of z multiplied by its probability density function p(q), over all possible values of q:

$$E[z] = \int zp(q)dq$$

Since z = Aq, we can substitute this expression for z in the above equation to obtain:

$$E[z] = \int (Aq)p(q)dq$$

We can factor out the constant matrix A from the integral, since it does not depend on q:

$$E[z] = A \int qp(q)dq$$

The integral $\int qp(q)dq$ is simply the expected value of q, denoted by E[q]:

$$E[z] = AE[q]$$

Therefore, we have shown that the expected value of the random variable z, which is defined as the matrix product of a constant matrix A and a multi-variate random variable q, is equal to the matrix product of A and the expected value of q.

1.2.2 2

Using the definition of Cov:

$$Cov[z] = E[(z - E[z])(z - E[z])^T]$$

First, we expand the product $(z - E[z])(z - E[z])^T$:

$$(z - E[z])(z - E[z])^T = (Aq - AE[q])(Aq - AE[q])^T = A(q - E[q])(q - E[q])^T A^T$$

where we have used the fact that A is a constant matrix and can be factored out of the expression.

Next, we substitute this expression into the formula for Cov[z]:

$$Cov[z] = E[A(q - E[q])(q - E[q])^T A^T]$$

Using the definition of covariance, we can write this expression as:

$$Cov[z] = AE[(q - E[q])(q - E[q])^T]A^T$$

where$E[(q-E[q])(q-E[q])^T]$ is the covariance matrix of q, denoted by Cov[q].

Therefore, we have:

$$Cov[z] = ACov[q]A^T$$

which is the desired expression for the covariance matrix of z in terms of the covariance matrix of q and the constant matrix A.

1.3 c

From the code, results of z are:

$$\mu_z = \begin{bmatrix} 5 \\ 10 \end{bmatrix}$$
$$\Sigma_z = \begin{bmatrix} 2.3 & 4 \\ 4 & 8 \end{bmatrix}$$

And the figure is:

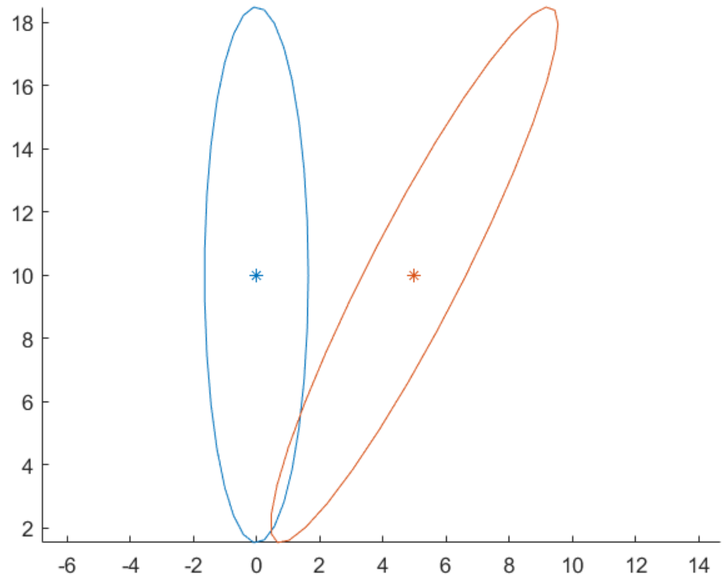


Figure 1: Covariance of Z and q

1.3.1 For mean

The matrix A shifts the mean of q by a factor of 0.5 in the x-direction, resulting in a mean of z that is shifted by 5 in the x-direction and unchanged in the y-direction.

1.3.2 For covariance

The correlation coefficient of p is 0 , while after transformation it turns to 4/(sqrt(2.3)*sqrt(8))=0.9325. It is also shown in the picture that the axis of the ellipse leans an angle after transformation.

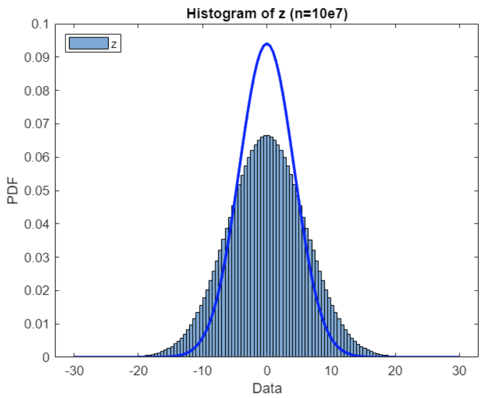
1.3.3 Traced back

According to the A matrix, this arises because the off-diagonal element of A reflects the fact that the x-component of z (z1) depends on both the x-component (q1) and the y-component (q2) of q. As a result, the transformation with A introduces a correlation between the individual components of z that was not present in q.

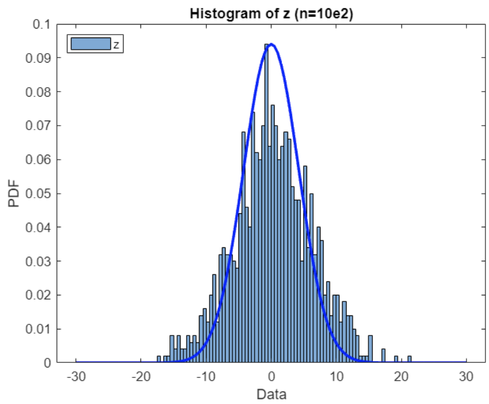
2 Transformation of random variables

2.1 a

It asks to use `approxGaussianTransform` method to draw histogram. Here are two graphs using different sample points:



(a) Number of points = 10e7



(b) Number of points = 10e2

Figure 2: Two sets of points

2.1.1 Does the histogram and two pdfs match well?

To read the covariance from a PDF curve, find the value of the axis for the peak of the curve and then determine the width of the curve at the half-maximum point, then the width is $2 \times \sigma$.

From figure 2, we can see both the mean and the covariance match, so my answer is YES.

2.1.2 From what you see in the figure, what conclusions can you draw regarding the properties of p(z) and the different approximations?

- The analytical and numerical approximations of the transformed Gaussian distribution are very similar in shape, with both pdfs having a peak around zero and a similar spread. This indicates that the numerical approximation using approxGaussianTransform is accurate and provides a good estimate of the true distribution.

- Linear transformations preserve the Gaussian nature of the original distribution(Compare with b get this)

2.1.3 How does the number of samples used in the approximation affect the result?

If the number of samples used in the approximation is small, then the approximation may not be accurate enough to capture the full shape of the probability distribution of the non-linear function. As the number of samples used in the approximation increases, the accuracy of the approximation should improve.

2.2 b

$$E[z] = E[x^3] = \int x^3 p(x) dx$$

Since x is a normal random variable with mean 0 and variance 2, its probability density function is given by:

$$p(x) = (1/sqrt(2 * pi * 2)) * exp(-x^2/4)$$

Substituting this into the above equation and evaluating the integral, we get:

$$E[z] = E[x^3] = 0$$

So, the mean of z is 0.

Next, we can find the covariance of z as follows:

$$Cov[z] = E[(z - E[z])^2] = E[z^2]$$

To find $E[z^2]$, we can use the following formula:

$$E[z^2] = \int z^2 p(z) dz$$

where $p(z)$ is the probability density function of z.

Substituting $z = x^3$ and using the change of variables formula, we get:

$$E[z^2] = E[x^6] = \int x^6 p(x)dx$$

To evaluate this integral, we can use the fact that $x^6 = (x^2)^3$and apply the same technique as before. We get:

$$E[z^2] = E[x^6] = 3 * Var[x]^3 = 3 * 2^3 = 24$$

Therefore, the covariance of z is:

$$Cov[z] = E[z^2] - E[z]^2 = 24 - 0^2 = 24$$

So, the mean of z is 0 and the covariance of z is 24.

Draw the picture for PDF and `approxGaussianTransform`:

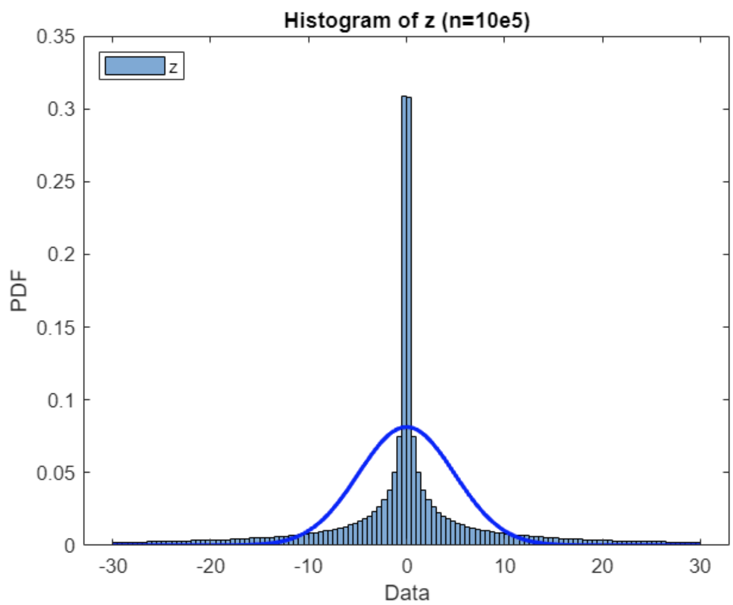


Figure 3: Histogram and PDF curve for b

From the figure, we can see the mean matches while the covariance does not match at all.

2.3 c

Linear transformations preserve the Gaussian nature of the original distribution while the non-linear transformation may result in a transformed distribution that is not Gaussian, even if the original distribution is Gaussian. This can make it more difficult to estimate the resulting distribution using a Gaussian approximation.

So it is reasonable to use `approxGaussianTransform` method when facing a linear transformation of Gaussian distribution while shouldn't believe the result of `approxGaussianTransform` method facing non-linear transformation

3 Understanding the conditional density

3.1 a

Since y is a linear combination of a known function h(x) and a normal random variable r, it follows that y is also a normal random variable with mean $\mu_y = E[h(x)]$ and variance $\sigma_y^2 = Var[h(x)] + \sigma_r^2$. The proof of this statement is as follows:

We know that h(x) is a deterministic function of x, so its mean and variance can be expressed as:

$$\mu_h = E[h(x)]$$

$$\sigma_h^2 = Var[h(x)]$$

Using the properties of expected values and variances, we can find the mean and variance of y as follows:

$$\mu_y = E[y] = E[h(x) + r] = E[h(x)] + E[r] = \mu_h + 0 = \mu_h$$

$$\sigma_y^2 = Var[y] = Var[h(x) + r] = Var[h(x)] + Var[r] + 2Cov[h(x), r]$$

Since r is a zero-mean normal random variable, we know that its variance is σ_r^2, and its covariance with h(x) is zero. Therefore, we can simplify the above expression as:

$$\sigma_y^2 = Var[h(x)] + \sigma_r^2$$

Thus, we have shown that y is a normal random variable with mean $\mu_y = E[h(x)]$ and variance $\sigma_y^2 = Var[h(x)] + \sigma_r^2$. Therefore, we can fully describe the distribution of y.

3.2 b

4 MMSE and MAP estimators

4.1 a

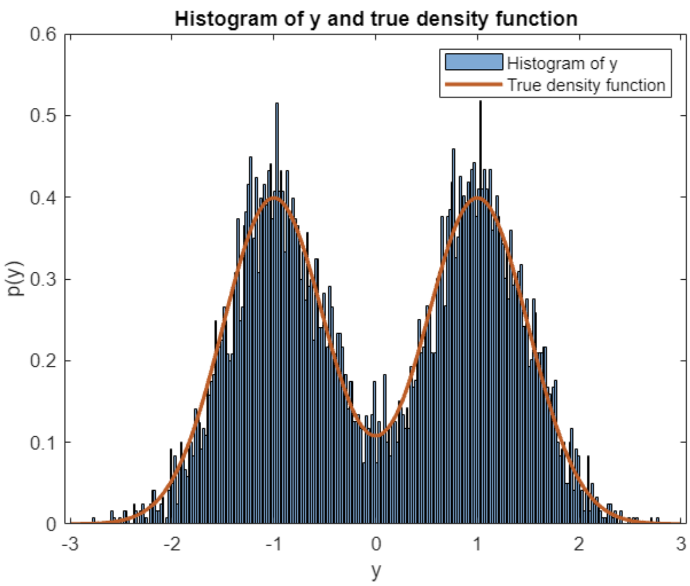


Figure 4: Sample 3000 times

Since θ is equally likely to be -1 or 1, the histogram of y should show two normal distributions with means -1 and 1 and equal variances $\sigma^2 = 0.25$. The overall shape of the histogram should be a bimodal distribution, with two peaks at -1 and 1 and equal weights. This bimodal distribution represents a mixture of two normal distributions, each with a different mean and the same variance.

4.2 b

Note: In my homework, solving b is after c. So I used the conclusions in c to prove b while.

In this problem, we are given that $y = \theta + w$, where $w \sim N(0, 0.25)$ is a normally distributed noise term with mean 0 and variance 0.25. Therefore, the probability density function of y given theta is given by:

$$p(y|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\theta)^2}{2\sigma^2}\right)$$

From c, we get:

$$p(y) = 0.5 \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-1)^2}{2\sigma^2}\right) + 0.5 \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y+1)^2}{2\sigma^2}\right)$$

Substituting $y = 0.7$ and $sigma^2 = 0.25$, we obtain:

```
q = @(theta,y) 1/sqrt(2*pi*sigma2)*exp(-(y-theta)^2/(2*sigma2));
q(1,0.7)
ans =

    0.6664


f = @(y) 0.5*1/sqrt(2*pi*sigma2)*exp(-(y-1)^2/(2*sigma2))+0.5*1/sqrt(2*pi*sigma2)*exp(-(y+1)^2/(2*sigma2));
f(0.7)

ans =

    0.3345
```

so:

$p(\theta = 1|y = 0.7) = \frac{p(y=0.7|\theta=1)p(\theta=1)}{p(y=0.7)} = \frac{0.6664 \times 0.5}{0.3345} = 0.9961$

$$p(\theta = -1|y = 0.7) = 0.0039$$

So I guess θ =1.

4.3 c

To prove that $p(y)$ is a mixture of two normal distributions with means -1 and 1 and the same variance σ^2, we can use the law of total probability.

In this problem, we can partition the sample space of y into two mutually exclusive events: $y = \theta + w$ where $\theta = 1$ and $\theta = -1$, where $w \sim \mathcal{N}(0, \sigma^2)$ is a normally distributed noise term with mean 0 and variance σ^2.

Using the law of total probability, we can express $p(y)$ as a mixture of the two normal distributions as follows:

$$p(y) = p(y|\theta = 1)p(\theta = 1) + p(y|\theta = -1)p(\theta = -1)$$

where $p(y|\theta = 1)$ is the probability density function of the normal distribution with mean 1 and variance σ^2, and $p(y|\theta = -1)$ is the probability density function of the normal distribution with mean -1 and variance σ^2.

Substituting the expressions for the two probability density functions, we obtain:

$$p(y) = 0.5 \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-1)^2}{2\sigma^2}\right) + 0.5 \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y+1)^2}{2\sigma^2}\right)$$

Therefore, $p(y)$ is a mixture of two normal distributions with means -1 and 1 and the same variance σ^2.

4.4 d

The Bayesian rule says:

$$p(\theta|y) = \frac{p(y|\theta)}{p(y)}$$

Already known $p(y)$ in question c and $p(y|\theta)$ in question b, so:

$$p(\theta|y) = \begin{cases} \frac{\frac{1}{2}\frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{1}{2\sigma^2}(y-1)^2\}}{\frac{1}{2}\frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{1}{2\sigma^2}(y-1)^2\}+\frac{1}{2}\frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{1}{2\sigma^2}(y+1)^2\}} & \text{if } \theta = 1 \\ \frac{\frac{1}{2}\frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{1}{2\sigma^2}(y+1)^2\}}{\frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{1}{2\sigma^2}(y-1)^2\}+\frac{1}{2}\frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{1}{2\sigma^2}(y+1)^2\}} & \text{if } \theta = -1 \end{cases}$$

$$= \begin{cases} \frac{\exp\{\frac{y}{\sigma^2}\}}{\exp\{\frac{y}{\sigma^2}\}+\exp\{-\frac{y}{\sigma^2}\}} & \text{if } \theta = 1 \\ \frac{\exp\{\frac{-y}{\sigma^2}\}}{\exp\{\frac{y}{\sigma^2}\}+\exp\{-\frac{y}{\sigma^2}\}} & \text{if } \theta = -1 \end{cases}$$

where $\sigma^2 = 0.25$

4.5 e

$$\begin{aligned} \hat{\theta}_{MMSE} &= \sum_\theta \theta \Pr\{\theta|y\}. \\ &= p(\theta = 1|y) - p(\theta = -1|y) \\ &= \frac{\exp\frac{y}{\sigma^2} - \exp-\frac{y}{\sigma^2}}{\exp\frac{y}{\sigma^2} + \exp-\frac{y}{\sigma^2}} \\ &= \frac{2\sinh\left(\frac{y}{\sigma^2}\right)}{2\cosh\left(\frac{y}{\sigma^2}\right)} = \tanh\left(\frac{y}{\sigma^2}\right) = tanh(4y) \end{aligned}$$

4.6 f

$$\begin{aligned} \hat{\theta}_{MAP} &= \arg\max_{\theta=\pm1} \pi_y(\theta) \\ &= \begin{cases} 1 & \text{if } \frac{\exp\frac{y}{\sigma^2}}{\exp\frac{y}{\sigma^2}+\exp-\frac{y}{\sigma^2}} \geq \frac{\exp-\frac{y}{\sigma^2}}{\exp\frac{y}{\sigma^2}+\exp-\frac{y}{\sigma^2}} \\ -1 & \text{if } \frac{\exp\frac{y}{\sigma^2}}{\exp\frac{y}{\sigma^2}+\exp-\frac{y}{\sigma^2}} < \frac{\exp-\frac{y}{\sigma^2}}{\exp\frac{y}{\sigma^2}+\exp-\frac{y}{\sigma^2}} \end{cases} \\ &= \begin{cases} 1 & \text{if } y \geq 0 \\ -1 & \text{if } y \leq 0 \end{cases} \end{aligned}$$

4.7 g

4.7.1 1

In 4b), we made the guess that θ is 1. Meanwhile, for the specific value of $y = 0.7$, both the MMSE estimator and the MAP estimator would predict that θ is 1. In this case, our guess in 4b coincides with both the MMSE and MAP estimators.

4.7.2 2

In this problem, the MMSE and MAP estimators for θ are different. The MMSE estimator is given by $\hat{\theta}_{MMSE} = \tanh(4y)$, while the MAP estimator is given by $\hat{\theta}_{MAP} = \mathrm{sgn}(y)$.

The MMSE estimator minimizes the expected squared error between the estimated value of θ and the true value, while the MAP estimator maximizes the posterior probability of θ given the observation y. In this case, the MMSE estimator and the MAP estimator make different decisions when y is close to 0.

If y is positive, both the MMSE and MAP estimators predict that θ is 1. If y is negative, the MMSE estimator predicts that θ is -1, while the MAP estimator predicts that θ is 1. This is because the MMSE estimator takes into account the entire posterior distribution of θ, while the MAP estimator only considers the most probable value of θ.