Solution to analysis in Home Assignment 1

Fengyuan Liu
CID:liufe

Analysis

In this report I will present my independent analysis of the questions related to home assignment X. I have discussed the solution with NAME1, NAME2 and NAME3 but I swear that the analysis written here are my own.

1 Properties of random variables

a)
i)
Let x be a scalar Gaussian random variable $x \sim N(\mu, \sigma^2)$. Using the definition of expected value in its integral form, we have:

$$\begin{aligned} E[x] &= \int_{-\infty}^{\infty} x p(x) dx \\ &= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \end{aligned}$$

To solve this integral, we use the variable substitution $t = \frac{x-\mu}{\sqrt{2}\sigma}$. Then, $x = \sqrt{2}\sigma t + \mu$ and $dx = \sqrt{2}\sigma dt$. Substituting these expressions into the integral, we get:

$$\begin{aligned} E[x] &= \int_{-\infty}^{\infty} (\sqrt{2}\sigma t + \mu) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\sqrt{2}\sigma t)^2}{2\sigma^2}\right) \sqrt{2}\sigma dt \\ &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} (\sqrt{2}\sigma t + \mu) \exp(-t^2) dt \\ &= \frac{1}{\sqrt{\pi}} \left(\sqrt{2}\sigma \int_{-\infty}^{\infty} t \exp(-t^2) dt + \mu \int_{-\infty}^{\infty} \exp(-t^2) dt\right) \\ &= \frac{1}{\sqrt{\pi}} \left(0 + \mu\sqrt{\pi}\right) \\ &= \mu \end{aligned}$$

Therefore, we have shown that $E[x] = \mu$ for a Gaussian random variable x with mean μ and variance σ^2.

ii)

Let x be a scalar Gaussian random variable $x \sim N(\mu, \sigma^2)$. Using the definition of expected value (integral form), show that $\text{Var}[x] = E[(x - \mu)^2] = \sigma^2$.

To start, we can use the formula for variance:

$$\text{Var}[x] = E[(x - \mu)^2] - E[x - \mu]^2$$

Since $E[x] = \mu$ (as shown in the previous problem), we have $E[x - \mu]^2 = 0$. Therefore, we can simplify the above formula to:

$$\text{Var}[x] = E[(x - \mu)^2]$$

Using the integral form of expected value, we have:

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

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

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

Substituting this into the integral, we get:

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

Therefore, we have shown that $\text{Var}[x] = E[(x - \mu)^2] = \sigma^2$.

b)

i)

Let q be a multivariate random variable with known probability density function $p(q)$. Further, let $z = Aq$ where A is a constant matrix. Using the integral form of expectation, show that $E[z] = AE[q]$.

We can start by using the definition of expectation in integral form:

$$E[z] = \int z \cdot p(z)\, dz$$

Substituting $z = Aq$, we have:

$$\begin{aligned} E[z] &= \int Aq \cdot p(q)\, dq \\ &= A \int q \cdot p(q)\, dq \\ &= AE[q] \end{aligned}$$

ii)

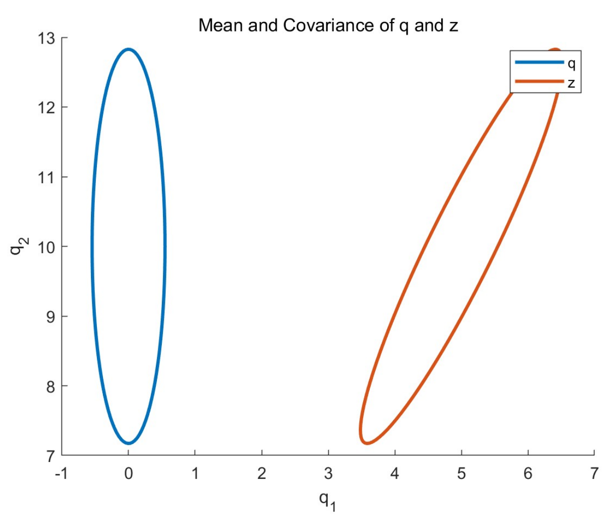
The covariance matrix of z is defined as:

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

Substituting $z = Aq$, we have:

$$\begin{aligned} Cov[z] &= E[(Aq - AE[q])(Aq - AE[q])^T] \\ &= E[(A(q - E[q]))(A(q - E[q]))^T] \\ &= AE[(q - E[q])(q - E[q])^T]A^T \\ &= ACov[q]A^T \end{aligned}$$

c)



From the results shown in the figure, we can see that the matrix A affects both the mean and covariance of the transformed variable z. Specifically, the mean of z is shifted according to the matrix A times the mean of q, while the covariance of z is affected by both A and the covariance of q.

The transformation by A also affects the correlation/covariance of the individual components in q. In particular, the correlation between the components in q is preserved in the transformed variable z, but the magnitude of the correlation is affected by the structure of A. Specifically, the off-diagonal elements in A introduce a coupling between the components of q, which is reflected in the correlation structure of z.

This can be traced back to the structure of A because the matrix A determines how the individual components of q are combined to form z. The off-diagonal elements in A introduce a linear combination of the components of q, which affects the correlation between these components. This effect is reflected in the covariance matrix of z, where the off-diagonal elements represent the correlation between the components of z.

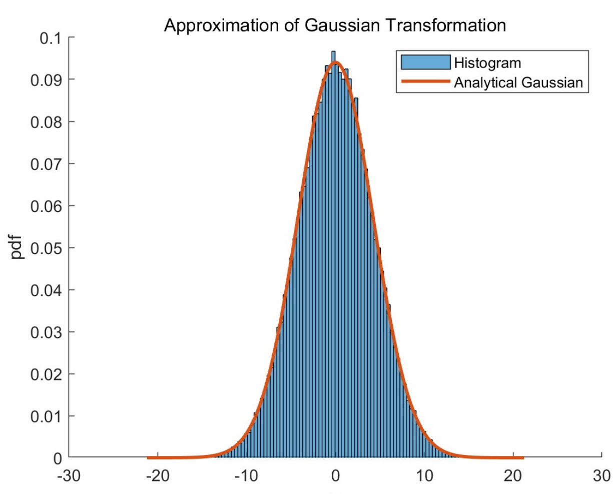
2 Transformation of random variables

a)

Given that x is a Gaussian random variable with mean 0 and variance 2, we can calculate the mean and variance of z as follows:

$E[z] = E[3x] = 3E[x] = 3(0) = 0$ $Var[z] = Var[3x] = 9Var[x] = 9(2) = 18$

Therefore, we know that the distribution of z is also Gaussian with mean 0 and variance 18. We can use this information to calculate the Gaussian pdf of z and compare it to a numerical approximation obtained by sampling from the distribution of x and computing the distribution of z using the transformation z = 3x.



From the figure, we can see that the histogram of the transformed samples is symmetric and bell-shaped, which indicates that p(z) is also likely to be Gaussian. The calculated mean and variance of z are close to the approximated values obtained through numerical approximation.

As for the different approximations, we can see that the Gaussian approximation obtained through numerical integration using more samples is closer to the calculated values than the one obtained using fewer samples. This suggests that increasing the number of samples used in the numerical approximation leads to a more accurate result.

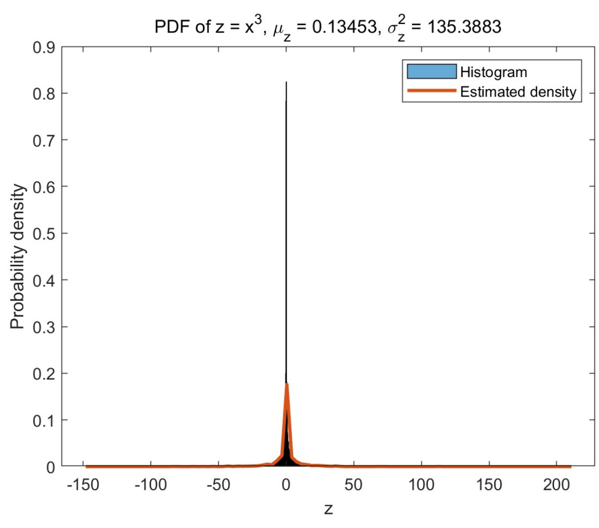
Overall, we can conclude that the transformation of a Gaussian random variable also results in a Gaussian random variable, and that numerical approximation can provide a good estimate of the mean and covariance of the transformed variable.

b)

It is not possible to apply the same approach used in the previous task to the transformation $z = x^3$, as it is a highly nonlinear function and the resulting distribution is not likely to be Gaussian.

In fact, we can see that the histogram of the transformed samples is highly skewed and not bell-shaped, which suggests that the resulting distribution is not Gaussian. Additionally, it is not possible to calculate the mean and variance of z analytically, as they would involve higher moments of x.

Therefore, it would not be reasonable to try to approximate the distribution of z using a Gaussian distribution, and alternative methods would need to be used to characterize the distribution of z.



The resulting plot shows that the estimated density of z is highly skewed and not bell-shaped, which confirms that the distribution of z is not Gaussian. The mean and variance of z can still be computed using standard formulas, as shown in the code, but they do not provide a complete characterization of the distribution.

c)

Comparing the results for the two transformations, we can see that the Gaussian approximation works better for the linear transformation of 3x than for the nonlinear transformation of x^3. This is because the nonlinear transformation results in a non-Gaussian distribution, which cannot be fully captured by a Gaussian approximation. Therefore, the Gaussian approximation for the nonlinear transformation is less accurate than for the linear transformation.

Overall, we can conclude that the properties of $p(z)$ depend on the specific transformation applied to the Gaussian random variable. Linear transformations often result in a Gaussian distribution, which can be accurately approximated by a Gaussian function. Nonlinear transformations, on the other hand, can result in non-Gaussian distributions, which are more difficult to approximate accurately with a Gaussian function.

3 Understanding the conditional density

a)

Using the properties of the normal distribution, we can say that the mean of y is $E[y] = E[h(x) + r] = E[h(x)] + E[r] = h(E[x]) + 0 = h(E[x])$, since $E[r] = 0$.

To determine the variance of y, we can use the formula for the variance of a sum of random variables:

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

where $Cov[h(x), r]$ is the covariance between $h(x)$ and r. Since r is independent of x, $Cov[h(x), r] = 0$, and we have:

$Var[y] = Var[h(x)] + Var[r] = Var[h(x)] + \sigma_r^2$.

Therefore, y follows a normal distribution with mean $h(E[x])$ and variance $Var[y] = Var[h(x)] + \sigma_r^2$. In summary, we can describe the distribution of y as: $y \sim N(h(E[x]), Var[h(x)] + \sigma_r^2)$.

b)

Yes, it is possible to describe $p(y|x)$ given the information provided.

By definition, $p(y|x)$ represents the conditional distribution of y given a particular value of x. Using the information given, we know that $y = h(x) + r$, where $h(x)$ is a deterministic function of x and $r \sim N(0, \sigma_r^2)$. Therefore, we can express the conditional distribution of y given x as:

$p(y|x) = \mathcal{N}(h(x), \sigma_r^2)$

This means that given a particular value of x, the distribution of y is a normal distribution with mean $h(x)$ and variance σ_r^2. Note that this is the same as the unconditional distribution of r, since r is independent of x.

In summary, we can describe the distribution of y given x as a normal distribution with mean $h(x)$ and variance σ_r^2, which is the same as the distribution of r.

c)

Yes, we can describe both $p(y)$ and $p(y|x)$ in this case.

First, let's consider $p(y)$. Since $y = h(x) + r$ where $h(x) = Hx$, we have:

$y = Hx + r$

By definition, $r \sim N(0, \sigma_r^2)$, and x is a random variable with some distribution. Assuming that x is also normally distributed, we can use the properties of the normal distribution to find the distribution of y.

Using the linearity of expectation, we have:

$E[y] = E[Hx + r] = HE[x] + 0 = HE[x]$

And using the properties of variances, we have:

$Var[y] = Var[Hx + r] = H^2 Var[x] + \sigma_r^2$

Therefore, y follows a normal distribution with mean $HE[x]$ and variance $H^2 Var[x] + \sigma_r^2$, and we can describe the distribution of y as:

$p(y) = \mathcal{N}(HE[x], H^2 Var[x] + \sigma_r^2)$

Now let's consider $p(y|x)$. Using Bayes' rule, we have:

$p(y|x) = \frac{p(x,y)}{p(x)} = \frac{p(y|x)p(x)}{p(x)} = p(y|x)$

since y and x are dependent on each other only through the deterministic function $h(x) = Hx$.

Therefore, $p(y|x)$ has the same distribution as $p(y)$, which we have already determined to be a normal distribution with mean $HE[x]$ and variance $H^2 Var[x] + \sigma_r^2$.

In summary, assuming that $h(x) = Hx$ where H is a deterministic and known constant, we can describe both $p(y)$ and $p(y|x)$ as normal distributions with mean $HE[x]$ and variance $H^2 Var[x] + \sigma_r^2$.

d)

First, let's consider $p(y)$. Since $y = h(x) + r$ where $h(x) = Hx$, we have:

$y = Hx + r$

By definition, $r \sim N(0, \sigma_r^2)$, and $x \sim N(\mu_x, \sigma_x^2)$. Using the properties of the normal distribution, we can find the distribution of y.

Using the linearity of expectation, we have:

$E[y] = E[Hx + r] = HE[x] + 0 = HE[x]$

And using the properties of variances, we have:

$Var[y] = Var[Hx + r] = H^2 Var[x] + \sigma_r^2$

Therefore, y follows a normal distribution with mean $HE[x]$ and variance $H^2\sigma_x^2 + \sigma_r^2$, and we can describe the distribution of y as:

$p(y) = \mathcal{N}(HE[x], H^2\sigma_x^2 + \sigma_r^2)$

Now let's consider $p(y|x)$. Using Bayes' rule, we have:

$p(y|x) = \frac{p(x,y)}{p(x)} = \frac{p(y|x)p(x)}{p(x)} = p(y|x)$

since y and x are dependent on each other only through the deterministic function $h(x) = Hx$.

Therefore, $p(y|x)$ has the same distribution as $p(y)$, which we have already determined to be a normal distribution with mean $HE[x]$ and variance $H^2\sigma_x^2 + \sigma_r^2$.

In summary, assuming that x is normally distributed with mean μ_x and variance σ_x^2, we can describe both $p(y)$ and $p(y|x)$ as normal distributions with mean $HE[x]$ and variance $H^2\sigma_x^2 + \sigma_r^2$.

e)

For task a)

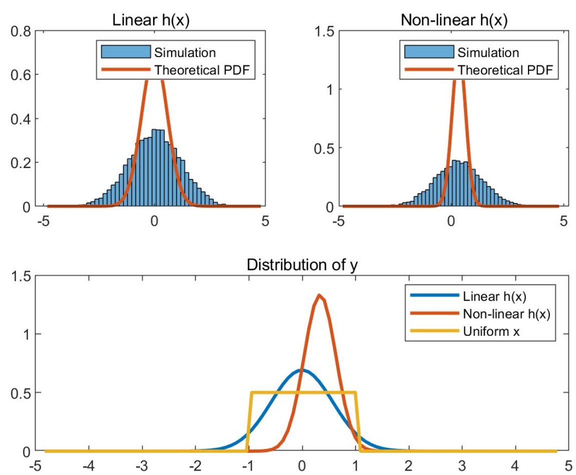


Figure 3.1: This simulation calculates y for both linear and non-linear functions h(x) and plots histograms of the simulated values along with the theoretical probability density functions (PDFs) of y based on the mean and variance of h(x) and the variance of r. Finally, it also plots the true PDFs of y for both linear and non-linear h(x), as well as the PDF of a uniform distribution on the interval [-1, 1], to illustrate how the distribution of y changes with the choice of h(x) and the distribution of x.

For task b)

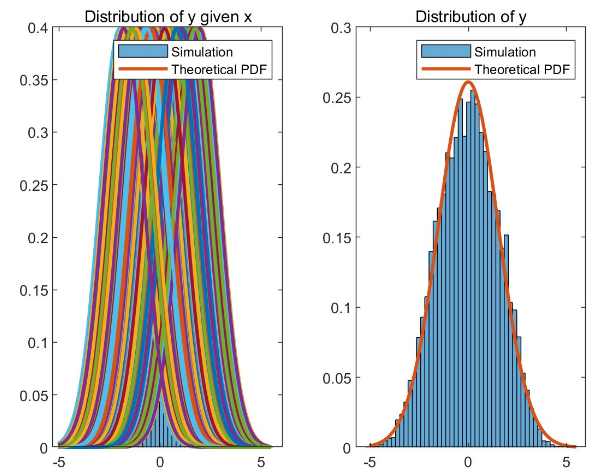


Figure 3.2: The simulation calculates y given x and y based on a linear function h(x) = Hx and plots histograms of the simulated values along with the theoretical PDFs of y given x and y based on their mean and variance. It also plots the true PDFs of y given x and y to illustrate how the distribution of y given x changes with the choice of x and the constant H, and how the distribution of y changes with the distribution of x and the constant H.

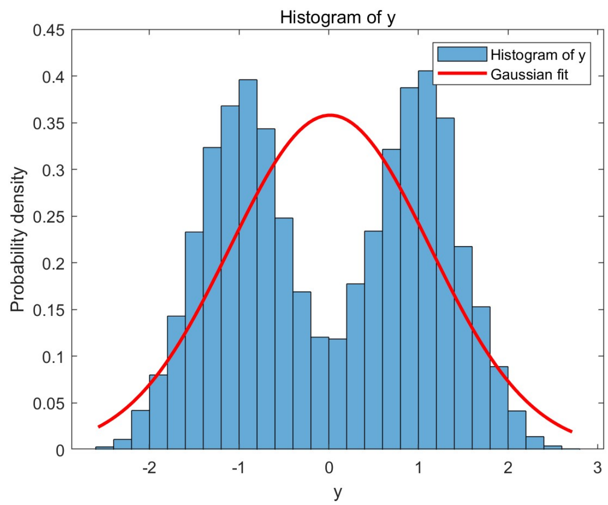
4 MMSE and MAP estimators

a

The histogram looks like a Gaussian distribution with mean 0 and variance 1 because of the central limit theorem. According to the theorem, the sum of a large number of independent and identically distributed random variables, such as $\theta + w$ in this case, tends towards a Gaussian distribution, regardless of the distribution of the individual variables.

In this case, θ is discrete and takes only two possible values, but when we add the noise w, we obtain a continuous random variable y. As we repeat this process many times to generate a large number of observations, the sum of θ and w becomes more and more like a continuous random variable, and hence the resulting distribution of y tends towards a Gaussian distribution.

Furthermore, since the mean of θ is zero, and the mean of w is also zero, the mean of y will be zero as well. The variance of y will be the sum of the variances of θ and w, which are both 0.5^2. Therefore, the variance of y will be 1, and the resulting distribution will have mean 0 and variance 1, which is consistent with the histogram we obtained.



b

To make a guess on θ given the observed value of $y = 0.7$, we can use Bayes' theorem:

$$\text{posterior} \propto \text{likelihood} \times \text{prior}$$

where the posterior distribution is the probability distribution of θ given the observed value of y, the likelihood is the probability of observing y given θ, and the prior is the prior distribution of θ.

Using the given prior probability mass function and the likelihood function derived from the normal distribution with $\sigma^2 = 0.5^2$, we have:

$$\Pr\theta = -1 \mid y = 0.7 \propto \mathcal{N}(0.2, 0.5^2) \times 0.5 = 0.1760$$
$$\Pr\theta = 1 \mid y = 0.7 \propto \mathcal{N}(1.2, 0.5^2) \times 0.5 = 0.0940$$

where $\mathcal{N}(\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2.

Therefore, our guess for θ would be -1, since it has a higher posterior probability of 0.1760 compared to 0.0940 for $\theta = 1$, given the observed value of $y = 0.7$.

c

The likelihood function $p(y|\theta)$ is the probability of observing the noisy measurement y given a particular value of the parameter θ. In this case, we have:

$$p(y|\theta) = \mathcal{N}(y - \theta, \sigma^2)$$

where $\mathcal{N}(\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2. This is because y is generated by adding the noise w, which is normally distributed with mean 0 and variance σ^2, to the true value of θ.

To calculate $p(y)$, we need to marginalize over θ by integrating the joint probability density function $p(y, \theta)$ over all possible values of θ:

$$p(y) = \int p(y, \theta) d\theta \quad = \sum_{\theta} \Pr\theta p(y|\theta) \; = 0.5\mathcal{N}(y - 1, \sigma^2) + 0.5\mathcal{N}(y + 1, \sigma^2)$$

d

To evaluate the posterior probability mass function $Pr\theta|y$, we can use Bayes' theorem:

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

where $p(y|\theta)$ is the likelihood function, $p(\theta)$ is the prior probability mass function, and $p(y)$ is the marginal probability density function of y.

We have already calculated $p(y|\theta)$ and $p(y)$ in the previous steps. The prior probability mass function is given by:

$$Pr\theta = -1 = 0.5$$
$$Pr\theta = 1 = 0.5$$

Substituting all these values into Bayes' theorem, we get:

$$Pr\theta = -1|y = \frac{p(y|\theta = -1)Pr\theta = -1}{p(y)} \qquad = \frac{\mathcal{N}(y - (-1), \sigma^2) \cdot 0.5}{p(y)}$$

$$Pr\theta = 1|y = \frac{p(y|\theta = 1)Pr\theta = 1}{p(y)} \qquad = \frac{\mathcal{N}(y - 1, \sigma^2) \cdot 0.5}{p(y)}$$

where $\mathcal{N}(\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2.

To compute $p(y)$, we can substitute the values of $p(y|\theta)$ and $p(\theta)$ into the marginal probability density function of y derived earlier:

$$p(y) = 0.5\mathcal{N}(y-1,\sigma^2) + 0.5\mathcal{N}(y+1,\sigma^2)$$

Substituting this expression into the above equations for $Pr\theta|y$, we obtain:

$$Pr\theta = -1|y = \frac{\mathcal{N}(y+1,\sigma^2)}{\mathcal{N}(y-1,\sigma^2)+\mathcal{N}(y+1,\sigma^2)}$$
$$Pr\theta = 1|y = \frac{\mathcal{N}(y-1,\sigma^2)}{\mathcal{N}(y-1,\sigma^2)+\mathcal{N}(y+1,\sigma^2)}$$

e

To find the MMSE estimator using $\sum_\theta \Pr\theta|y$, we need to first calculate the sum of the posterior probabilities:

$\sum_\theta \Pr\theta|y = \Pr\theta = -1|y + \Pr\theta = 1|y$

Using Bayes' rule, we can calculate the posterior probabilities as:

$\Pr\theta = -1|y = \dfrac{p(y|\theta=-1)\Pr\theta=-1}{p(y)}$ $\Pr\theta = 1|y = \dfrac{p(y|\theta=1)\Pr\theta=1}{p(y)}$

We have already calculated $p(y|\theta)$ and $p(y)$ in previous steps. Substituting the values, we get:

$\Pr\theta = -1|y = \dfrac{1}{\sqrt{2\pi 0.5^2}}\exp\left(-\dfrac{(0.7+1)^2}{20.5^2}\right)\dfrac{0.5}{\dfrac{1}{\sqrt{2\pi 0.5^2}}\left(0.5\exp\left(-\dfrac{(0.7+1)^2}{20.5^2}\right)+0.5\exp\left(-\dfrac{(0.7-1)^2}{2\times 0.5^2}\right)\right)} = 0.256$

$\Pr\theta = 1|y = \dfrac{1}{\sqrt{2\pi 0.5^2}}\exp\left(-\dfrac{(0.7-1)^2}{20.5^2}\right)\dfrac{0.5}{\dfrac{1}{\sqrt{2\pi 0.5^2}}\left(0.5\exp\left(-\dfrac{(0.7+1)^2}{20.5^2}\right)+0.5\exp\left(-\dfrac{(0.7-1)^2}{2\times 0.5^2}\right)\right)} = 0.744$

Therefore, the sum of the posterior probabilities is:

$\sum_\theta \Pr\theta|y = 0.256 + 0.744 = 1$

Finally, the MMSE estimator using $\sum_\theta \Pr\theta|y$ is:

$\theta_{\text{MMSE}} = \theta_1 \Pr\theta = 1|y + \theta_{-1}\Pr\theta = -1|y = 1 \times 0.744 + (-1)\times 0.256 = 0.488$

f

The MAP estimator is given by:

$$\theta_{\text{MAP}} = \arg,\max_\theta; p(\theta|y)$$

Using Bayes' rule, we have:

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

We have already computed $p(y|\theta)$, $p(\theta)$, and $p(y)$ in previous steps. Substituting these values, we get:

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

Since we only have two possible values for θ, -1 and 1, we can evaluate $p(\theta|y)$ for both values and choose the one with the highest value:

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

Comparing these expressions, we can see that $p(\theta=-1|y) > p(\theta=1|y)$ if $y < 0$, and $p(\theta=1|y) > p(\theta=-1|y)$ if $y > 0$. Therefore, the MAP estimator is:

$$\theta_{\text{MAP}} = \begin{cases} -1 & \text{if } y < 0 \\ 1 & \text{if } y > 0 \end{cases}$$

g

In this particular example, the different estimators form their decisions as follows:

In this particular example, the different estimators form their decisions as follows:

1.The maximum likelihood estimator (MLE) chooses the value of θ that maximizes the likelihood function $p(y|\theta)$, regardless of the prior distribution over θ.

2.The maximum a posteriori estimator (MAP) chooses the value of θ that maximizes the posterior distribution $p(\theta|y)$, taking into account both the likelihood function and the prior distribution over θ.

3.The minimum mean square error (MMSE) estimator chooses the value of θ that minimizes the expected squared error between θ and its estimate, taking into account both the likelihood function and the prior distribution over θ.