Solution to analysis in Home Assignment 1

Abdulrahman Hameshli
CID: hameshli

Contents

Analysis

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

1 Properties of random variables

1.1 a

We start with $E[x]$, integrating x with $p(x)$. After expanding $p(x)$, we factor out terms independent of x. Introducing $t = \frac{x-\mu}{\sigma}$, we simplify the integral, where $dx = \sigma dt$. This simplification leads to two integrals, one evaluates to 0 due to symmetry, the other to $\mu\sqrt{2\pi}$. Thus, $E[x] = \mu$.

Math: Gaussian distribution density function is:

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

This gives us that the expected value can be calculated as:

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

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

This can be expanded to:

$$\begin{aligned}\mathrm{Var}[x] &= E[x^2 - 2x\mu + \mu^2] \\ &= E[x^2] - 2\mu E[x] + \mu^2 \\ &= E[x^2] - 2\mu^2 + \mu^2 \\ &= E[x^2] - \mu^2\end{aligned}$$

$$\begin{aligned}\mathrm{Var}[x] &= \int_{-\infty}^{\infty} x^2 p(x)dx - 2\mu^2 + \mu^2 \\ &= \int_{-\infty}^{\infty} x^2 p(x)dx - \mu^2 \\ \text{Letting } t &= \frac{x-\mu}{\sigma}, \text{ we can rewrite the integral as:} \\ \mathrm{Var}[x] &= \sigma^2 \int_{-\infty}^{\infty} (\sigma t + \mu)^2 p(t)dt - \mu^2 \\ &= \sigma^2 \left(\int_{-\infty}^{\infty} t^2 p(t)dt + 2\mu \int_{-\infty}^{\infty} tp(t)dt + \mu^2 \int_{-\infty}^{\infty} p(t)dt - \mu^2 \right) \\ &= \sigma^2 \left(1 + 0 + \mu^2 - \mu^2\right) \\ &= \sigma^2\end{aligned}$$

1.2 b)

The expected value of a random variable z is computed by integrating the product of the variable z and its probability density function $p(z)$. Through a change of variable $z = Aq$, where A is a constant, the expression is simplified, ultimately yielding $AE[q]$, where $E[q]$ represents the expected value of q.

$$Z(q) = Aq$$

$$E[z] = \int zp(z)dz$$

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

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

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

The covariance of a random variable z is shown to be equivalent to the covariance of q scaled by the constant A. This is achieved by expanding the expression for covariance using the definition and properties of covariance matrices.

$$\mathrm{Cov}[z] = \mathrm{Cov}[Aq] = E[(Aq - E[Aq])(Aq - E[Aq])^\top]$$

$$\text{Cov}[z] = E[(Aq - AE[q])(Aq - AE[q])^\top]$$
$$\text{Cov}[z] = E[A(q - E[q])(q - E[q])^\top A^\top]$$
$$\text{Cov}[z] = AE[(q - E[q])(q - E[q])^\top]A^\top$$
$$\text{Cov}[z] = A\text{Cov}[q]A^\top$$

1.3 c)

$$\mu = \begin{bmatrix} 0 \\ 10 \end{bmatrix}$$
$$A = \begin{bmatrix} 1 & 0.5 \\ 0 & 1 \end{bmatrix}$$
$$\sigma = \begin{bmatrix} 0.3 & 0 \\ 0 & 8 \end{bmatrix}$$
$$Z = A \cdot q$$

The expected value can be calculated as:

$$E[z] = A \cdot E[q] =$$
$$E[z] = \begin{bmatrix} 1 & 0.5 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 10 \end{bmatrix} = \begin{bmatrix} 5 \\ 10 \end{bmatrix}$$

To calculate the variance of Z:

$$\text{Var}[z] = A \cdot x \cdot A^\top$$

This gives us:

$$p(z) = \mathcal{N}\left(z; \begin{bmatrix} 5 \\ 10 \end{bmatrix}, \begin{bmatrix} 2.3 & 4.0 \\ 4.0 & 8.0 \end{bmatrix}\right)$$

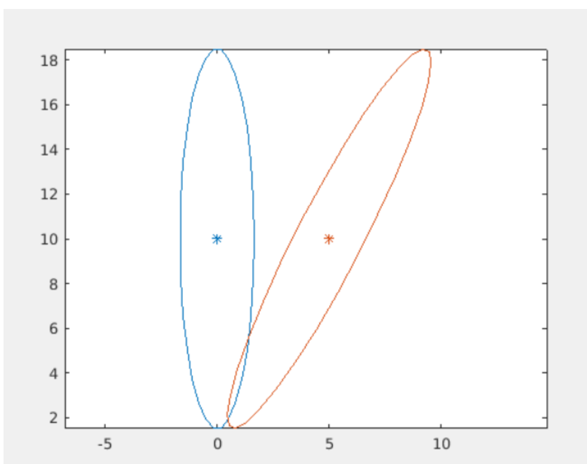


Figure 1.1: Mean and variance of X (blue) and Z (orange)

In Figure 1.1, we observe that Z demonstrates dependencies between q_1 and q_2, suggesting that an increase in q_1 coincides with an increase in q_2. This relationship is further supported by the covariance matrix for Z, which shows a covariance of 4 between the two variables, indicating their dependence. The cross-covariance is induced by the transformation governed by the matrix A, specifically by the entry 0.5 in the first row of the second column.

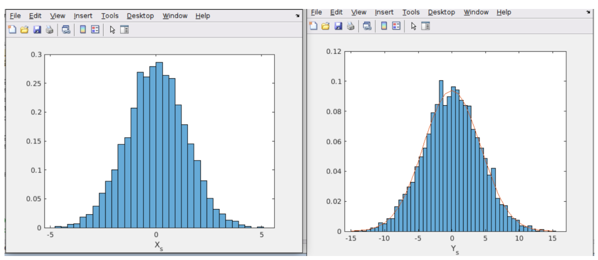
2 2

2.1 a

$$f = 3 * x$$

The result can be seen bellow. We use 5000 samples to estimate the PDF in both cases, and it's clear that they look quite similar as expected. The number of iterations significantly impacts the approximation and can greatly influence its accuracy. Increasing the amount of data leads to a more precise estimation.

$$x \sim \mathcal{N}(0, 2)$$
$$\hat{z}_{\text{analytical}} \sim \mathcal{N}(0, 18)$$
$$\hat{z}_{\text{approximate}} \sim \mathcal{N}(-0.0158, 18.0417)$$

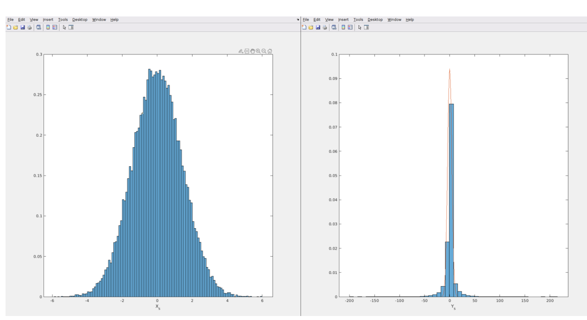


2.2 b

$$f = x^3$$

The function is non-linear which means it is hard to determine the analytical expression for mean and variance but we can still approximate it using fucnition "approxGaussianTransform". The number of samples is = 500000.

$$x \sim \mathcal{N}(0, 2)$$
$$\hat{z}_{\text{approximate}} \sim \mathcal{N}(-0.0323, 118.6174)$$



2.3 c

In this particular case, we observe that there is no analytical solution for the non-linear function, and the mean remains at 0, indicating no impact on the mean. However, if the function includes a constant, for example, $x^3 + C$, the mean will shift to the constant C. On the other hand, the non-linear function does have a significant influence on the variance.

3 3

3.1 a

The variance of h(x) is equal to 0 since it is a deterministic function.

The given funciton is:

$$y = h(x) + r$$

where r $\sim \mathcal{N}(0, \sigma^2)$
$E(y) = E(h(x)) + E(r)$ $E(h(x)) = h(x)$ and $E(r) = 0$. Therefore $E(y) = h(x)$.

The variance is

$$\text{var}(y) = E((y - E(y))^2) = E(r^2) = \text{var}(r) = \sigma^2$$

However if h(x) has a constant then the mean of y will converge to the constant of h(x) and variance would be the σ_r Therefor it is possible to know what distribution y has.

let

$$h(x) = x^2 + C$$

then

$$y\ \mathcal{N}(C, \sigma_r^2)$$

3.2 b

Yes.

$$y\ \mathcal{N}(C, \sigma_r^2)$$

3.3 c

Yes.

$$y\ \mathcal{N}(HX, \sigma_r^2)$$

3.4 d

3.5 e

4 4

4.1 a

The distribution is bimodal, or we could say that y consists of two overlapping normal distributions. y depends on θ, where θ is either 1 or -1. This implies that the most frequently occurring values for y should be 1 or -1, hence there are two peaks and the mean is located at 0.

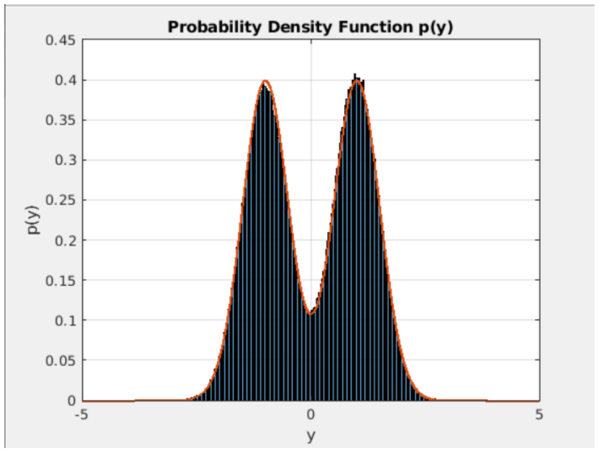


Figure 4.1: The red line is the true distribution without randomizing. And the data is shown with randomizing the theta and w $N(0, 0.5^2)$

4.2 b

Theta is set to $\theta = 1$ because it is more reasonable that $\theta = 1$ when $y = 0.7$. There is a higher probability that $\theta = 1$ when $y = 0.7$. This relationship can be expressed as $y = \theta + w = 1 \pm w$.

Certainly. When $y = 0.7$, if $\theta = 1$, y is expected to be around 1 with a small positive deviation w, which is quite likely. However, if $\theta = -1$, y is expected to be around -1 with a small negative deviation w, which is less likely. Therefore, given $y = 0.7$, it's more reasonable to infer that $\theta = 1$.

4.3 c

To find the conditional probability density function $p(y|\theta)$, we consider the noise model $y = \theta + w$, where $w \sim \mathcal{N}(0, \sigma^2)$.

Given θ with discrete prior probabilities, we derive distributions for y conditioned on $\theta = -1$ and $\theta = 1$.

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

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

The conditional PDF $p(y|\theta)$ is a mixture of these distributions:

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

4.4 d

Using Bayes' rule, we can calculate the posterior probability $\Pr(\theta|y)$ by combining the prior probability $\Pr(\theta)$, likelihood $\Pr(y|\theta)$, and evidence $\Pr(y)$. Given that θ can only take the values $\theta = -1$ or $\theta = 1$, we evaluate the posterior separately for each value of θ.

Below is the distribution when $y = 0.7$, which is reasonable. There is more probability that θ is 1 when $y = 0.7$.

```
Posterior probability for theta = -1 given y: 0.003684
Posterior probability for theta = 1 given y: 0.996316
```

4.5 e

$$\theta_{MMSE} = \sum_{\theta} \theta * Pr(|y) = 1 * \text{posterior}_{\theta_1} + (-1) * \text{posterior}_{\theta_{-1}}$$

where :

$$\text{posterior}_{\theta_{-1}} = \frac{\text{likelihood}_{y|\theta_{-1}} \times \text{prior}_{\theta_{-1}}}{\text{total\_probability}_y}$$

$$\text{posterior}_{\theta_1} = \frac{\text{likelihood}_{y|\theta_1} \times \text{prior}_{\theta_1}}{\text{total\_probability}_y}$$

Observing y = 0.7 the mean of theta given y: 0.992632

(see code section, Q4,e))

4.6 f

4.7 g

5 Code:

```
%% Q1
%Here we give parameters for a Gaussian density. The parameter mu is the mean, and P is the covariance matrix.
mu = [0; 10];
P = [0.3 0;0 8];

muz = [5 ;10];
Pz=[2.3 4;4 8];

%Call your function.
xy = sigmaEllipse2D(mu, P);
z = sigmaEllipse2D(muz,Pz);
%Now plot the generated points. You should see an elongated ellipse stretching from the top left corner to the bottom right.
figure(1);
h1 = plot(xy(1,:), xy(2,:));
hold on
h2=plot(z(1,:), z(2,:));
%Set the scale of x and y axis to be the same. This should be done if the two variables are in the same domain, e.g. both are measured in meters.
axis equal
hold on
%Also plot a star where the mean is, and make it have the same color as the ellipse.
plot(mu(1), mu(2), '*', 'color', h1.Color);
plot(muz(1), muz(2), '*', 'color', h2.Color);



%% Q2

f_1 = @(x)[3*x];
f_2 = @(x)(x.^3 + 300);
[mu_y, Sigma_y, y_s ,x_s]= approxGaussianTransform(0,2,f_2,500000);

mu_y
Sigma_y

figure(1)
histogram(y_s);
xlabel("Y_s")

figure(2)
histogram(x_s);
xlabel("X_s")

% % True  values:
% disp('the true values is');
% [mu_y, Sigma_y] = affineGaussianTransform(0,1, 3, 0)


%% Q3

```

```matlab
%% Q4
% Part a)
clc; clear all;

% Generate random noise
w = normrnd(0, 0.5, 1000);

% Define possible values for theta and their probabilities
values = [-1, 1];
probs = [0.5, 0.5];

% Sample theta values according to probabilities
theta = randsample(values, 1000, true, probs);

% Generate observed data
y = theta + w;

% Plot histogram of observed data
histogram(y, 'Normalization', 'pdf');
hold on;

% Define parameters for the true distribution of y
sigma2 = 0.5^2;
y_range = linspace(-5, 5, 10000);

% Calculate the true probability density function of y
p_y = 0.5 * (1/sqrt(2*pi*sigma2)) * exp(-(y_range - 1).^2 / (2*sigma2)) + ...
      0.5 * (1/sqrt(2*pi*sigma2)) * exp(-(y_range + 1).^2 / (2*sigma2));

% Plot the true distribution
plot(y_range, p_y, 'LineWidth', 2);
xlabel('y');
ylabel('p(y)');
title('Probability Density Function p(y)');
grid on;

%% Parts c and d)
% Given variables
sigma_sq = 0.5^2;                % Variance of the noise w
prior_theta_minus_1 = 0.5;       % Prior probability for theta = -1
prior_theta_1 = 0.5;             % Prior probability for theta = 1

% Observed data y
observed_y = 0;

% Calculate likelihoods p(y | theta = -1) and p(y | theta = 1)
likelihood_y_given_theta_minus_1 = (1 / sqrt(2*pi*sigma_sq)) * exp(-(
    observed_y + 1).^2 / (2*sigma_sq));
likelihood_y_given_theta_1 = (1 / sqrt(2*pi*sigma_sq)) * exp(-(observed_y - 1)
```

```
    .^2 / (2*sigma_sq));

% Calculate total probability of observing y: p(y)
total_probability_y = 0.5 * likelihood_y_given_theta_minus_1 + 0.5 * likelihood_y_given_theta_1;

% Calculate posterior probabilities for theta = -1 and theta = 1
posterior_theta_minus_1 = (likelihood_y_given_theta_minus_1 * prior_theta_minus_1) / total_probability_y;
posterior_theta_1 = (likelihood_y_given_theta_1 * prior_theta_1) / total_probability_y;

% Display the results
fprintf('Posterior probability for theta = -1 given y: %f\n', posterior_theta_minus_1);
fprintf('Posterior probability for theta = 1 given y: %f\n', posterior_theta_1);

%% Part e)
% Calculate the mean of theta using posterior probabilities
theta_mean = posterior_theta_1 * 1 + posterior_theta_minus_1 * (-1);

% Display the mean of theta
fprintf('Mean of theta given y: %f\n', theta_mean);
```