Multivariate Gaussian Exercise

HY-673, 2023-2024, Spring

Exercise 2 (2022-2023 Assignment #1)

(a) Let X_1, X_2 be two *dependent* Gaussian r.v.s with $\mathbb{E}[X_i] = \mu_i$, and $\mathrm{Var}(X_i) = \sigma_i^2$, for $i = 1, 2$, and $\mathbb{E}[(X_1 - \mu_1)(X_2 - \mu_2)] = \sigma_{12}$. Assume that the sum $Y = X_1 + X_2$ is also a Gaussian r.v. and compute the PDF of Y.

Solution:

In order to compute the PDF of Y, where $Y = X_1 + X_2$, $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$, which we know is Gaussian, we can take advantage of the fact that a Gaussian distribution is uniquely characterized by its mean and variance. We can compute the mean μ_Y by using the linearity of expectation:

$$\mu_Y = \mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2] = \mu_1 + \mu_2. \tag{1}$$

Additionally, we know that the variance of the sum of two r.v.s equals the sums of the variances of those r.v.s, plus two times their covariance. To prove it, we can use the definition of variance:

$$\begin{aligned}
\sigma_Y^2 &= \mathbb{E}\left[(X_1 + X_2 - \mu_Y)^2\right] &(2)\\
&= \mathbb{E}\left[(X_1 + X_2 - \mu_1 - \mu_2)^2\right] &(3)\\
&= \mathbb{E}\left[(X_1 - \mu_1)^2 + (X_2 - \mu_2)^2 + 2(X_1 - \mu_1)(X_2 - \mu_2)\right] &(4)\\
&= \mathbb{E}\left[(X_1 - \mu_1)^2\right] + \mathbb{E}\left[(X_2 - \mu_2)^2\right] + 2\mathbb{E}\left[(X_1 - \mu_1)(X_2 - \mu_2)\right] &(5)\\
&= \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2\mathrm{Cov}(X_1, X_2) &(6)\\
&= \sigma_1^2 + \sigma_2^2 + 2\sigma_{12}, &(7)
\end{aligned}$$

because it is given that $\mathrm{Var}(X_i) = \sigma_i^2$, $i = 1, 2$, and $\mathbb{E}[(X_1 - \mu_1)(X_2 - \mu_2)] = \sigma_{12}$. Hence, the answer is $Y \sim \mathcal{N}\left(\mu_1 + \mu_2,\ \sigma_1^2 + \sigma_2^2 + 2\sigma_{12}\right)$.

We observe that if we denote the mean vector and covariance matrix for the joint distribution X_1, X_2 as:

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix},\ \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2, \end{bmatrix}, \tag{8}$$

we have that (i) the mean μ_Y is the sum of the two elements of the mean vector, and (ii) the variance σ_Y^2 is the sum of all elements of the covariance matrix. We will use this in the following question.

(b) Compute $p(x_1 + x_2 | x_3)$, given that $p(x_1, x_2, x_3) \equiv \mathcal{N}(\mu, \Sigma)$, $\mu \in \mathbb{R}^3$, $\Sigma \in \mathbb{R}^{3\times 3}$.

Solution:

Without loss of generality, let us denote the elements of the mean vector and covariance matrix for the triplet X_1, X_2, X_3 as follows:

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}, \; \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix}. \tag{9}$$

We can first compute the **joint** conditional distribution $\mathbf{p(x_1, x_2|x_3)}$ by using the theory we know from the second lecture of the course (multivariate normal distribution section). In short, our theory says that:

$$p(x_A|x_B) = N\left(\mu', \Sigma'\right), \tag{10}$$

where the mean μ' we are looking for is given by the formula:

$$\mu' = x_A|\mu_A + \Sigma_{AB}\Sigma_{BB}^{-1}(x_A - \mu_B), \tag{11}$$

and the covariance matrix Σ' by:

$$\Sigma' = \Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA}. \tag{12}$$

So, in our case, we have $x_A = [x_1, x_2]^\intercal$, and $x_B = x_3$, i.e.:

$$\mu_A = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \; \mu_B = \mu_3, \tag{13}$$

$$\Sigma_{AA} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}, \; \Sigma_{BB} = \sigma_3^2, \tag{14}$$

$$\Sigma_{AB} = \begin{bmatrix} \sigma_{13} \\ \sigma_{23} \end{bmatrix}, \; \Sigma_{BA} = [\sigma_{13} \; \sigma_{23}]. \tag{15}$$

Therefore, we straightly plug in the above in eq. (11) to find the mean:

$$\mu' = \mu_{X_1, X_2|x_3} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \sigma_{13} \\ \sigma_{23} \end{bmatrix} \frac{1}{\sigma_3^2}(x_3 - \mu_3) = \begin{bmatrix} \mu_1 + \frac{\sigma_{13}}{\sigma_3^2}(x_3 - \mu_3) \\ \mu_2 + \frac{\sigma_{23}}{\sigma_3^2}(x_3 - \mu_3) \end{bmatrix} = \begin{bmatrix} \mu_{Y_1} \\ \mu_{Y_2} \end{bmatrix}, \tag{16}$$

and in eq. (12) to find the covariance matrix:

$$\Sigma' = \Sigma_{X_1, X_2|x_3} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} - \begin{bmatrix} \sigma_{13} \\ \sigma_{23} \end{bmatrix} \frac{1}{\sigma_3^2} \begin{bmatrix} \sigma_{31} & \sigma_{32} \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} - \frac{1}{\sigma_3^2} \begin{bmatrix} \sigma_{13}^2 & \sigma_{13}\sigma_{32} \\ \sigma_{23}\sigma_{31} & \sigma_{23}^2 \end{bmatrix} \tag{17}$$

$$= \begin{bmatrix} \sigma_1^2 - \frac{\sigma_{13}^2}{\sigma_3^2} & \sigma_{12} - \frac{\sigma_{13}\sigma_{32}}{\sigma_3^2} \\ \sigma_{21} - \frac{\sigma_{23}\sigma_{31}}{\sigma_3^2} & \sigma_2^2 - \frac{\sigma_{23}^2}{\sigma_3^2} \end{bmatrix} = \begin{bmatrix} \sigma_{Y_1} & \sigma_{Y_1, Y_2} \\ \sigma_{Y_2, Y_1} & \sigma_{Y_2} \end{bmatrix}. \tag{18}$$

Now, we know that since $p(x_1, x_2|x_3)$ follows a **multivariate** Gaussian distribution, then $p(x_1+x_2|x_3)$ should follow a **univariate** Gaussian distribution. Using our observation from the previous question of the exercise (final paragraph), we can find its mean and variance by summing the elements of the mean vector and covariance matrix respectively. Therefore, for the mean, it will be:

$$\mu_{X_1+X_2|x_3} = \mu_1 + \mu_2 + \frac{\sigma_{13} + \sigma_{23}}{\sigma_3^2}, \tag{19}$$

and for the variance:

$$\sigma^2_{X_1+X_2|x_3} = \sigma_1^2 + \sigma_2^2 + 2\sigma 12 - \frac{1}{\sigma_3^2}\left(\sigma_{13}^2 + \sigma_{23}^2 + 2\sigma_{13}\sigma_{23}\right). \tag{20}$$

Hence, we found that $p(x_1 + x_2|x_3)$ is a Gaussian PDF with mean $\mu_{X_1+X_2|x_3}$, and variance $\sigma^2_{X_1+X_2|x_3}$.

(c) Using $p(x_1, x_2|x_3)$ as an intermediate step, numerically validate through histograms comparisons the result in (b), for:

$$\mu = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \; \Sigma = \begin{bmatrix} 1 & -0.5 & 0.8 \\ -0.5 & 1 & -0.8 \\ 0.8 & -0.8 & 1 \end{bmatrix}. \tag{21}$$

Solution:

The code for the numerical evaluation is in `03d_multivariate_gaussian.ipynb`.

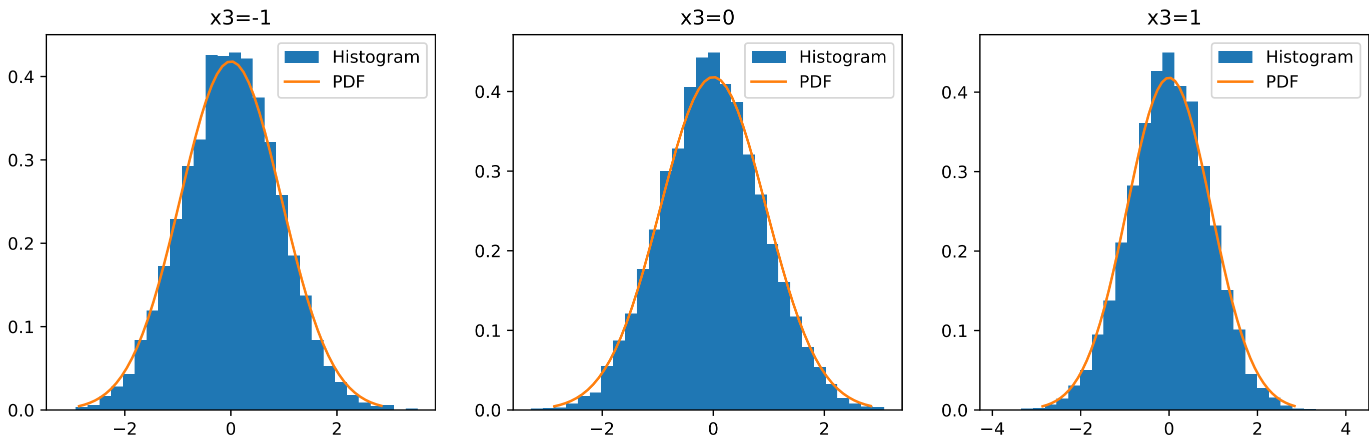

Figure 1: We get the same distribution no matter the value of x_3 due to the covariance matrix symmetries.

<u>Comments:</u> We are comparing the empirical distribution of Y and its PDF for $x_3 = -1, 0, 1$. The value of x_3 does not change the distribution of Y. This is due to the structure of the covariance matrix: X_1 and X_2 have opposite mean and are negative correlated with each other. Furthermore, they have the same degree of correlation with X_3. This implies that whatever value X_3 takes, it will "even out" when X_1 and X_2 are summed. This can also be verified by the fact that x_3 gets canceled out if we plug the mean and covariance values in eq. (19).

Bonus: Analytic proof for the sum of two Gaussian r.v.s

From theory, we know that (i) we can always express the sum of two **dependent** Gaussians, say, $Y = X_1 + X_2$, as the sum of two **independent** Gaussians, but, it is also a well-known fact that (ii) the sum of two independent Gaussians results in a single Gaussian. Combining these two facts together, we reach to the conclusion that the sum of two dependent Gaussians is also always a Gaussian, i.e., it always holds that $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ in our case. Here, we will prove why the second fact holds.

As we know, the PDF of $Y = X_1 + X_2,\ X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2),\ X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ can be computed with convolving the PDFs:

$$p_Y(y) = \int_{-\infty}^{\infty} p_{X_1}(x) p_{X_2}(y - x) dx. \tag{22}$$

However, this integral is long to calculate, and, therefore, prone to mistakes. A better alternative is to find the *characteristic function*[1] of the Gaussian distribution. This way, convolutions will become products, and our job will be much easier:

$$\begin{aligned}
\varphi_x(\omega) &= \int_{-\infty}^{+\infty} p_x(x) e^{j\omega x} dx && (23)\\
&= \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} e^{j\omega x} dx && (24)\\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\left(\frac{(x-\mu)^2}{2\sigma^2} - j\omega x\right)} dx && (25)\\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-\left(x^2 - 2x\mu + \mu^2 - 2\sigma^2 j\omega x\right)/2\sigma^2\right) dx && (26)\\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-\left((x - (\mu + \sigma^2 j\omega))^2 + 2\mu\sigma^2 j\omega + \sigma^4\omega^2\right)/2\sigma^2\right) dx && (27)\\
&= e^{-\omega\left(j\mu + \sigma^2\omega/2\right)} \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{\left((x-(\mu+\sigma^2 j\omega))^2\right)}{2\sigma^2}} dx && (28)\\
&= e^{-\omega\left(j\mu + \omega\sigma^2/2\right)}. && (29)
\end{aligned}$$

From the definition of the characteristic function, we have:

$$\mathbb{E}\left[e^{j(X_1+X_2)}\right] = \mathbb{E}\left[e^{jX_1}\right]\mathbb{E}\left[e^{jX_2}\right],\ X_1 \perp\!\!\!\perp X_2. \tag{30}$$

Therefore, the characteristic function of $Y = X_1 + X_2$, will be:

$$\varphi_y(\omega) = \varphi_{x_1}(\omega)\varphi_{x_2}(\omega) = e^{-\omega\left(j\mu_1 + \sigma_1^2\omega/2\right)} e^{-\omega\left(j\mu_2 + \sigma_2^2\omega/2\right)} = e^{-\omega(j(\mu_1+\mu_2) + \omega(\sigma_1^2+\sigma_2^2)/2)}, \tag{31}$$

which is the characteristic function of a Gaussian as seen by eq. 29, with mean $\mu_y = \mu_1 + \mu_2$, and variance $\sigma_y^2 = \sigma_1^2 + \sigma_2^2$.

[1] The characteristic function of a probability distribution can be thought of as a specific type of Fourier transform evaluated at $-\omega$ that is applied to probability distributions.
https://en.wikipedia.org/wiki/Characteristic_function_(probability_theory)