Statistical Pattern Recognition
Assignment 1

Ali Gholami

Department of Computer Engineering & Information Technology
Amirkabir University of Technology

http://ceit.aut.ac.ir/~aligholamee
aligholamee@aut.ac.ir

Abstract

This is an introductory assignment to the world of *Statistics* and *Probability* in the context of *Pattern Recognition*. We'll introduce some key concepts like *Probability Distribution Function, Cumulative Distribution Function, Probability Density Function, Probability Mass Function, Joint Probability Density Function, Joint Cumulative Density Function, Marginal Density* & more details as the probabilistic point of view. Furthermore, we'll review the concepts of *Expected Value, Variance, Standard Deviation, Covariance & Correlation of Random Variables(e.g. Random Vectors), Univariate & Multivariate Gaussian Distribution, Total Probability & Bayes Theorem, Geometric & Mahalanobis Distances, Central Limit Theorem, Independence & Correlation* as the statistics point of view. Also, a principal concept called *Linear Transformation* is discussed. The relationship between these fields is far more important than each separately.

Keywords. *PDF, PMF, JPDF, JPMF, CDF, JCDF, Covariance Matrix, Correlation Coefficient, Correlation, Variance, Expected Vector, Gaussian Distribution, Marginal Probability, Linear Tranformation, Eigenvector, Eigenvalue, Rank.*

1 Expectation & Variance

A random variable X has $E(X) = -4$ and $E(X^2) = 30$. Let $Y = -3X + 7$. Compute the following.

(a) $V(X)$

(b) $V(Y)$

(c) $E((X + 5)^2)$

(d) $E(Y^2)$

Solution

The main equation to calculate the *Variance* of a random variable X is given in 1.1.

$$V(X) = E[(X - E[X])^2] \tag{1.1}$$

Expanding the equation 1.1, we'll have the equation 1.2 using simple calculus.

$$V(X) = E[X^2 + E[X]^2 - 2XE[X]]$$
$$V(X) = E[X^2] + E[X]^2 - 2E[X]^2$$
$$V(X) = E[X^2] - E[X]^2 \tag{1.2}$$

(a) The equation 1.2 can be conducted directly to compute the *Variance*. Replacing the values from the problem description we get the following as result.

$$V(X) = E[X^2] - E[X]^2$$
$$V(X) = 30 - 16 = 14$$

(b) It is important to mention the *Linearity* of *Expectation* operator as formally described in 1.3.

$$E[aX + b] = aE[X] + b \tag{1.3}$$

Using property 1.3, we can write the *V(Y)* as

$$V(Y) = V(-3X + 7) = E[(-3X + 7)^2] - E[-3X + 7]^2$$
$$V(Y) = E[9X^2 + 49 - 42X] - E[-3X + 7]^2$$
$$V(Y) = 9E[X^2] + E[49] - 42E[X] - 9E[X]^2 - E[49]$$
$$V(Y) = 9 * 30 + 49 - 42 * (-4) - 9 * 30 - 49$$
$$V(Y) = 168$$

(c) Expanding the internals of the expectation, we'll get the following.

$$E[(X + 5)^2] = E[X^2 + 10X + 25]$$
$$E[(X + 5)^2] = E[X^2] + 10E[X] + E[25]$$
$$E[(X + 5)^2] = 30 + 10 * (-4) + 25$$
$$E[(X + 5)^2] = 15$$

(d) Same as above, we'll use 1.2 to get the following.

$$E[Y^2] = E[(3X + 7)^2]$$
$$E[Y^2] = E[9X^2 + 49 - 42X]$$
$$E[Y^2] = 487$$

2   Eigenvector & Eigenvalue

(a) Compute eigenvalues and eigenvectors of $A = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 9 & -5 \end{bmatrix}$ and compare your results

with Matlab outputs.

(b) A $2 * 2$ matrix A has $\lambda_1 = 2$ and $\lambda_2 = 5$, with corresponding eigenvectors $V_1 = \begin{bmatrix} 1 & 0 \end{bmatrix}^T$ and $V_2 = \begin{bmatrix} 1 & 1 \end{bmatrix}^T$. Find A.

Solution

(a) First of all, we'll find the *eigenvalues* using the *characteristic equation* given in 2.1.

$$|A - \lambda I| = 0 \tag{2.1}$$

Using this equation, all of the diagonal components of matrix A will be decremented by a λ term. The determinant of the resulting matrix will be as following.

$$A - \lambda I = \begin{bmatrix} 4 - \lambda & 0 & 0 \\ 0 & 2 - \lambda & 2 \\ 0 & 9 & -5 - \lambda \end{bmatrix}$$

$$|A - \lambda I| = (4 - \lambda)(\lambda^2 + 3\lambda - 28)$$

This will result in the following two roots for the λ term.

$$\boxed{\lambda_1 = 4} \quad \boxed{\lambda_2 = -7}$$

These are the *eigenvalues* of the given matrix A. We'll continue using the *Gaussian Elimination* technique to compute the *eigenvectors* of matrix A. According to the equation 2.2 we are looking for all possible vectors that can be substitute with vector X.

$$|A - \lambda I|X = 0 \tag{2.2}$$

Thus, for each *eigenvalue* determined in the previous computations, we'll find the proper *eigenvector* by converting the equation 2.2 to a *Row Echelon Form* and solving the resulting linear system by *Back Substitution*. Using the computed *eigenvalues*, we'll have 2.3 and 2.4

$$A - 4I = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -2 & 2 \\ 0 & 9 & -9 \end{bmatrix} \tag{2.3}$$

$$A + 7I = \begin{bmatrix} 11 & 0 & 0 \\ 0 & 9 & 2 \\ 0 & 9 & 2 \end{bmatrix} \tag{2.4}$$

3

Now we can find the proper X for each *eigenvalue*, using *augmented* version of 2.3 and 2.4. We'll have 2.5 and 2.6 as result.

$$E_1 = (A - 4I \mid 0) = \begin{bmatrix} 0 & 0 & 0 & \vline & 0 \\ 0 & -2 & 2 & \vline & 0 \\ 0 & 9 & -9 & \vline & 0 \end{bmatrix} \tag{2.5}$$

$$E_2 = (A + 7I \mid 0) = \begin{bmatrix} 11 & 0 & 0 & \vline & 0 \\ 0 & 9 & 2 & \vline & 0 \\ 0 & 9 & 2 & \vline & 0 \end{bmatrix} \tag{2.6}$$

The above matrices will be converted to the *Row Echelon Form* below using *Row Operations*.

$$E_1 = \begin{bmatrix} 0 & 9 & -9 & \vline & 0 \\ 0 & -2 & 2 & \vline & 0 \\ 0 & 9 & -9 & \vline & 0 \end{bmatrix}$$

$$E_1 = \begin{bmatrix} 0 & 9 & -9 & \vline & 0 \\ 0 & -2 & 2 & \vline & 0 \\ 0 & 0 & 0 & \vline & 0 \end{bmatrix}$$

$$E_1 = \begin{bmatrix} 0 & 1 & -1 & \vline & 0 \\ 0 & -2 & 2 & \vline & 0 \\ 0 & 0 & 0 & \vline & 0 \end{bmatrix}$$

$$E_1 = \begin{bmatrix} 0 & 1 & -1 & \vline & 0 \\ 0 & 0 & 0 & \vline & 0 \\ 0 & 0 & 0 & \vline & 0 \end{bmatrix}$$

The resulting equation for E_1 will be as following.

$$(0)(X_1) + (1)(X_2) - (1)(X_3) = 0$$

Thus, we'll have the following *eigenvector* and *eigenvalue*.

$$X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} 0 \\ X_2 \\ X_2 \end{bmatrix}$$

$$X = X_2 \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \lambda = 4$$

for any real number $X_1 \neq 0$. Calculating the *eigenvector* using the same procedure, we'll get the following results.

$$E_2 = (A + 7I \mid 0) = \begin{bmatrix} 11 & 0 & 0 & \vline & 0 \\ 0 & 9 & 2 & \vline & 0 \\ 0 & 9 & 2 & \vline & 0 \end{bmatrix} \tag{2.7}$$

Using *Row Operations* we'll have:

$$E_2 = \left[\begin{array}{ccc|c} 11 & 0 & 0 & 0 \\ 0 & 9 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Since E_2 is in the *Row Echelon Form*, we can infer the equation for E_2 as:

$$(11)(X_1) = 0$$

$$(9)(X_2) + 2(X_3) = 0$$

For the *eigenvector* we'll have:

$$X = X_2 \begin{bmatrix} 0 \\ 1 \\ \frac{-9}{2} \end{bmatrix}, \quad \lambda = -7$$

for any real number $X_2 \neq 0$.

Python Outputs

```
Eigenvalues:  [-7.  4.  4.]
Eigenvectors:  [
[ 0.          0.          1.         ]
[ 0.21693046 -0.70710678  0.         ]
[-0.97618706 -0.70710678  0.         ]
]
```

Comparing Results

The outputs from *linalg.eig* function of *numpy* library depicts that the *eigenvalues* for this matrix are -7 and 4. The eigenvector has been calculated for an specific X_1 and X_2. In this case we can obtained the outputs by replacing $X_2 = -0.70710$, $X_2 = 1$ and $X_2 = 0.2169$. Note that having $\lambda = 4$ gives us $y = z$ also which means that we can write its *eigenvector* form like so.

$$X = X_2 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \lambda = 4$$

(b) Since A is a $2 * 2$ matrix, we have:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

5

Given *eigenvectors* and *eigenvalues* in the problem description, we can conduct the *Characteristic Equation* presented in 2.2 and get the following results.

$$(A - 2I)X = \begin{bmatrix} a - 2 & b \\ c & d - 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$(A - 5I)X = \begin{bmatrix} a - 5 & b \\ c & d - 5 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Simply solving these equations will give us the following results.

$$\boxed{a = 2} \quad \boxed{b = 3} \quad \boxed{c = 0} \quad \boxed{d = 5}$$

Finally, we'll rewrite the matrix A replacing the new parameters in it.

$$A = \begin{bmatrix} 2 & 3 \\ 0 & 5 \end{bmatrix}$$

3   Probability Density Function

Let
$$f(x, y) = \begin{cases} c(x + y^2) & if \ \ 0 < x < 1 \ ; \ 0 < y < 1 \\ 0 & elsewhere. \end{cases}$$

Find each of the following.

(a) c

(b) $f_X(x)$

(c) $f_{X|Y}(y)$

(d) Are X and Y independent?

(e) What is the probability $Pr(X < \frac{1}{2}|Y = \frac{1}{2})$?

Solution

(a) The *Joint Probability Distribution* is given above. It is obvious that the total probability is always equal to 1; Hence we can integrate over the variables in the *JPDF* and that will give us the *Total Probability* which is equal to *1*.

$$\int\limits_0^1 \int\limits_0^1 f(x, y) \ \mathrm{d}x \ \mathrm{d}y = 1$$

Replacing the given equation we'll have the following.

$$\int\limits_0^1 \int\limits_0^1 c(x + y^2) \ \mathrm{d}x \ \mathrm{d}y = 1$$

$$\int\limits_0^1 (\frac{c}{2} + cy^2) \ \mathrm{d}y = 1$$

$$\frac{c}{2} + \frac{c}{3} = 1$$

$$\boxed{c = \frac{6}{5}}$$

(b) *Marginal Density* can easily be computed by integrating over the variable of not interest.

$$f_X(x) = \int_0^1 cx + cy^2 \ \mathrm{d}y$$

7

$$f_X(x) = \int_0^1 \frac{6x}{5} + \frac{6y^2}{5} \, dy$$

$$\boxed{f_X(x) = \frac{6x+2}{5}}$$

(c) While having the relation for the *Joint Probability* and the ease of finding *Marginal Density* of each of the variables, we can simply find the conditional probability using the *Bayes* theorem.

$$f_{X|Y}(x) = \frac{f(x,y)}{g(y)} = \frac{joint \ density \ of \ x \ \& \ y}{marginal \ density \ of \ y} \tag{3.1}$$

The $g(y)$ can be found by integrating the *JPDF* over the variable X.

$$g(y) = \int_0^1 \frac{6}{5}(x + y^2) \, dx$$

$$g(y) = \frac{6y^2}{5} + \frac{3}{2}$$

Finally, the $f_{X|Y}(x)$ will be:

$$f_{X|Y}(x) = \frac{\frac{6}{5}(x + y^2)}{\frac{6y^2}{5} + \frac{3}{2}}$$

(d) In the case the X and Y are independent, the conditional probability $f_{X|Y}(x)$ is equal to $f_X(x)$. In the previous section, we calculated the $f_{X|Y}(x)$ using *Bayes* theorem. We can simply find $f_X(x)$ from the *JPDF* by integrating over Y. If the results are equal:

$$f_{X|Y}(x) = f_X(x)$$

we say that X and Y are independent. Computing the integral of *JPDF* over Y we get:

$$f_X(x) = \frac{6x+2}{5}$$

So the variables X and Y are not *independent*.

(e) Given the conditional density introduced in 3.1, we'll have:

$$Pr(X < \frac{1}{2} \mid Y = \frac{1}{2}) = \frac{f(X < \frac{1}{2}, Y = \frac{1}{2})}{g(\frac{1}{2})}$$

$$Pr(X < \frac{1}{2} \mid Y = \frac{1}{2}) = \frac{\frac{3}{10}}{\frac{36}{20}}$$

$$Pr(X < \frac{1}{2} \mid Y = \frac{1}{2}) = \frac{1}{6}$$

8

4 Multivariate Gaussian Distribution

Recall the multivariate Gaussian for a vector $x \, \epsilon \, R^n$.

$$p(x \mid \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{\frac{-1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)} \tag{4.1}$$

(a) Let $y = v_i^T x$, where v_i is an eigenvector of Σ with an eigenvalue of λ_i. Find the probability density of $p(Y)$.

(b) Let X be a zero mean Gaussian random vector with an isotropic covariance ($\Sigma = I$). Let $Y = AX + B$. Compute the mean and variance of Y.

(c) (*Matlab*) Generate 500 random samples from a 2 dimensional Gaussian with an isotropic Σ using the *Matlab* command **randn**. Transform the data as above with $B = \begin{bmatrix} 0.5 \\ 1 \end{bmatrix}$, and $A = \begin{bmatrix} -5 & 5 \\ 1 & 1 \end{bmatrix}$. Plot the original and transformed points.

Solution

(a) The represented map in section a is a *Linear Map*. Applying a *Linear Map* to a random vector X with *mean* μ_X and *covariance matrix* Σ_X will give us a new random vector Y with the following properties.

$$\boxed{\Sigma_Y = v_i \Sigma v_i^T} \tag{4.2}$$

$$\boxed{\mu_Y = v_i \mu} \tag{4.3}$$

It is important to mention that the result of a *Linear Map* on a random vector X with *Gaussian* distribution will result in variable Y with *Gaussian* distribution; Hence we can simply represent the new density function for the variable Y using the *covariance matrix* and *mean* vector given 4.1, 4.2 and 4.3.

$$p(y \mid \mu_Y, \Sigma_Y) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma_Y|^{\frac{1}{2}}} e^{\frac{-1}{2}(y-\mu_Y)^T \Sigma_Y^{-1}(y-\mu_Y)} \tag{4.4}$$

Thus we will have the following probability distribution for random vector Y.

$$p(y \mid \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |v_i \Sigma v_i^T|^{\frac{1}{2}}} e^{\frac{-1}{2}(y-v_i\mu)^T (v_i \Sigma v_i^T)^{-1}(y-v_i\mu)}$$

SIDE NOTE: THE EIGENVECTOR OF A COVARIANCE MATRIX IS EQUAL TO PRINCIPAL COMPONENT. THE EIGENVECTOR WITH THE LARGEST EIGENVALUE IS ALONG WHICH THE DATASET HAS THE MAXIMUM VARIANCE. OUR INTEREST IS TO MAXIMIZE THE VARIANCE OF DATA IN THE NEW DIMENSIONAL SPACE AFTER APPLYING THE LINEAR MAP.

(b) A formal representation of the problem description is given below.

$$X \sim \mathcal{N}(\mu_X, \ \Sigma_X) = \mathcal{N}(0, \ I)$$

Note that the isotropic covariance will look like $C = \lambda I$ which in this special case $\lambda = 1$. Applying the linear map $y = Ax + B$ on X will have the following effect on the *mean* and *covariance matrix.*

$$Y \sim \mathcal{N}(A\mu_X + B, \ v_i \Sigma_X v_i^T)$$
$$Y \sim \mathcal{N}(B, I)$$

The new *mean* will be equal to B and the new *covariance matrix* will be equal to I because:

$$Cov(Y) = Cov(AX + B) = A \ \Sigma \ A^T = A \ I \ A^T = A \ A^T \ I = I$$

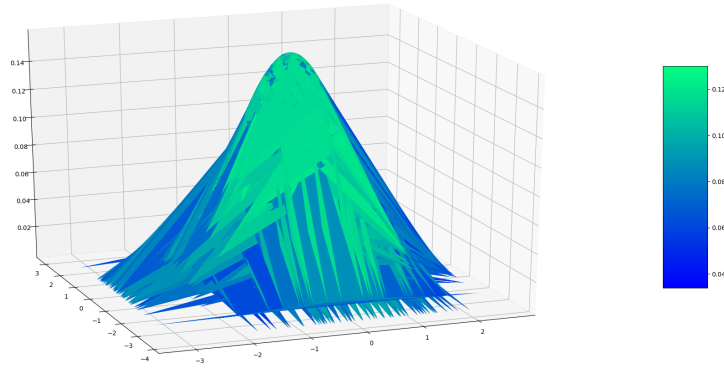(c) The output of the *Axes3D* of *Matplotlib* is given below.

Figure 4.1: 3D Plot of a 2D Gaussian Distribution with zero mean and isotropic covariance.

5   Correlation of Random Variables

Given $X_1, \ X_2, \ X_3, \ X_4$ as *uncorrelated* random variables each having *mean* 0 and *variance* 1, compute the correlations of:

(a) $X_1 + X_2$ and $X_2 + X_3$

(b) $X_1 + X_2$ and $X_3 + X_4$

Solution

Before getting into the details of answer, let's talk a bit about the *covariance* and *correlation* terms. They both represent the degree to which two random variables or sets of random variables tend deviate from their expected values. According to the *Wikipedia*, the *covariance* and *correlation* of two random variables X and Y is computed using 5.1 and 5.2 respectively.

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] \tag{5.1}$$

$$\rho_{XY} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \tag{5.2}$$

The difference between *covariance* and *correlation* is clear. *correlation* is a *dimensionless* measure and *covariance* is in *units*. Keeping this in mind, we can safely keep moving through the calculation of (a) and (b).

(a) Using the covariance equation for two random variables which is given in 5.3 we' ll get the following.

$$Cov(A, \ B) = E[AB] - E[A]E[B] \tag{5.3}$$

$$Cov(A, \ B) = E[X_1X_2 + X_1X_3 + X_2^2 + X_2X_3] - E[X_1 + X_2]E[X_2 + X_3]$$

$$Cov(A, \ B) = E[X_1X_2] + E[X_1X_3] + E[X_2^2] + E[X_2X_3]$$

$$-E[X_1]E[X_2] - E[X_1]E[X_3] - E[X_2]^2 - E[X_2]E[X_3]$$

Since the variables are *uncorrelated* pairwise, we have:

$$Cov(X_1, X_2) = 0$$

$$Cov(X_1, X_3) = 0$$

$$Cov(X_1, X_4) = 0$$

$$Cov(X_2, X_3) = 0$$

$$Cov(X_2, X_4) = 0$$

$$Cov(X_3, X_4) = 0$$

Therefore:

$$E[X_1X_2] = E[X_1]E[X_2]$$

$$E[X_1X_3] = E[X_1]E[X_3]$$

$$E[X_2X_3] = E[X_2]E[X_3]$$

Finally, the results for the $Cov(A, \ B)$ will be:

$$Cov(A, \ B) = E[X_2^2] - E[X_2]^2$$

which is the *variance* of X_2 according to the 1.2 and the value of that is equal to 1 according to the problem description. Thus, we'll have the following result for the correlation of $A = X_1 + X_2$ and $B = X_2 + X_3$.

$$\boxed{\rho_{AB} = 1}$$

Which represents that A and B have a *perfect positive* relationship; when one variable moves higher or lower, the other variable moves in the same direction with the same magnitude.

(b) Using the same terminology as previous section, we'll have the following computations to find the correlation of $A = X_1 + X_2$ and $B = X_3 + X_4$.

$$Cov(A, \ B) = E[X_1X_3 + X_1X_4 + X_2X_3 + X_2X_4] - E[X_1 + X_2]E[X_3 + X_4]$$

$$Cov(A, \ B) = E[X_1X_3] + E[X_1X_4] + E[X_2X_3] + E[X_2X_4]$$

$$-E[X_1]E[X_3] - E[X_1]E[X_4] - E[X_2]E[X_3] - E[X_2]E[X_4]$$

$$Cov(A, \ B) = 0$$

$$\boxed{\rho_{AB} = 0}$$

6   Independence & Expectation

If X and Y are *independent* and *identically distributed* with *mean* λ and *variance* σ^2, find $E[(X-Y)^2]$.

Solution

We have the following information about X and Y random variables:

$$E[X] = \lambda$$
$$E[Y] = \lambda$$
$$V[X] = \sigma^2$$
$$V[Y] = \sigma^2$$

Expanding the given equation for the X and Y we get:

$$E[(X-Y)^2] = E[X^2] + E[Y^2] - 2E[X]E[Y]$$

$$E[(X-Y)^2] = E[X^2] + E[Y^2] - 2\lambda^2 \tag{6.1}$$

Using the equation 1.2 we can simply find $E[X^2]$ and $E[Y^2]$:

$$E[X^2] = \sigma^2 + \lambda^2$$

$$E[Y^2] = \sigma^2 + \lambda^2$$

Replacing with materials in equation 6.1 we'll get:

$$\boxed{E[(X-Y)^2] = 2(\sigma^2 + \lambda^2) - 2\lambda^2 = 2\sigma^2}$$

7   Marginal & Conditional Density

Let $p(X)$ be $\mathcal{N}(M, \Sigma)$ with

$$X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad M = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} \ and \ \Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

Show that:

(a) $p(x_1) \sim \mathcal{N}_{x_1}(m_1, \ \sigma_1^2)$ (*a marginal density*)

(b) $p(x_1 \mid x_2) \sim \mathcal{N}_{x_1}(m_1 + \frac{\rho\sigma_1(x_2 - m_2)}{\sigma_2}, \ \sigma_1^2(1 - \rho^2))$ (*a conditional density*)

Solution

Since we'll need some *inverse matrix* computations, we need to review two theorems. The first theorem is about *inverse of a partitioned symmetric matrix* which divides every square matrix into 4 blocks. Correspondingly, the inverse matrix will be divided into 4 blocks and the inverse of these blocks can be calculated separately.

Theorem 1

Let $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^T & A_{22} \end{bmatrix}$, which demonstrates the division of every square matrix into 4 blocks. The inverse matrix $B = A^{-1}$ can also be divided into four blocks.

$$B = A^{-1} = \begin{bmatrix} B_{11} & B_{12} \\ B_{12}^T & B_{22} \end{bmatrix}$$

Now we can separately compute each of the components of inverse matrix.

$$B_{11} = (A_{11} - A_{12}A_{22}^{-1}A_{12}^T)^{-1} = A_{11}^{-1} + A_{11}^{-1}A_{12}(A_{22} - A_{12}^T A_{11}^{-1} A_{12})^{-1} A_{12}^T A_{11}^{-1}$$

$$B_{22} = (A_{22} - A_{12}^T A_{11}^{-1} A_{12})^{-1} = A_{22}^{-1} + A_{22}^{-1} A_{12}^T (A_{11} - A_{12} A_{22}^{-1} A_{12}^T)^{-1} A_{12} A_{22}^{-1}$$

$$B_{12} = -A_{22}^{-1} A_{12}^T (A_{11} - A_{12} A_{22}^{-1} A_{12}^T)^{-1}$$

$$B_{12}^T = -A_{11}^{-1} A_{12} (A_{22} - A_{12}^T A_{11}^{-1} A_{12})^{-1}$$

Theorem 2

Let $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$, which A is a partitioned symmetric matrix. We can compute the determinant of matrix A using the following formula:

$$\left| A \right| = \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} = \left| A_{11} \right| \left| A_{22} - A_{12}^T A_{11}^{-1} A_{12} \right| = \left| A_{22} \right| \left| A_{11} - A_{11} A_{22}^{-1} A_{12}^T \right|$$

Now let's derive the *multivariate Gaussian* for vector X. The *joint density* of X is:

$$p(x_1,\ x_2 \mid \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{\frac{-1}{2} D(x_1,\ x_2)} \tag{7.1}$$

Expanding the $D(x_1,\ x_2)$ as a matrix multiplication, we get:

$$D(x_1, x_2) = \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}^T \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}$$

The inverse matrix will be displayed with Σ^{ij} according to the conventions. Thus, by using *Theorem 1* we get:

$$D(x_1,\ x_2) = \begin{bmatrix} (x_1 - \mu_1)^T & (x_2 - \mu_2)^T \end{bmatrix} \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}$$

14

The result of the multiplication can be shown as following.

$$D(x_1,\ x_2) = (x_1 - \mu_1)^T \Sigma^{11} (x_1 - \mu_1) + 2(x_1 - \mu_1)^T \Sigma^{12}(x_2 - \mu_2) + (x_2 - \mu_2)^T \Sigma^{22}(x_2 - \mu_2)$$

We will replace the Σ^{11}, Σ^{12} and Σ^{22} with the result from *Theorem 2*.

$$D(x_1,\ x_2) = (x_1 - \mu_1)^T [\Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12})^{-1}\Sigma_{12}^T\Sigma_{11}^{-1}](x_1 - \mu_1)$$

$$-2(x_1 - \mu_1)^T[\Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12})^{-1}](x_2 - \mu_2)$$
$$+(x_2 - \mu_2)^T[(\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12})^{-1}](x_2 - \mu_2)$$

Applying multiplication on the first term we get:

$$D(x_1,\ x_2) = (x_1 - \mu_1)^T\Sigma_{11}^{-1}(x_1 - \mu_1)$$

$$+(x_1 - \mu_1)^T[\Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12})^{-1}\Sigma_{12}^T\Sigma_{11}^{-1}](x_1 - \mu_1)$$
$$-2(x_1 - \mu_1)^T[\Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12})^{-1}](x_2 - \mu_2)$$
$$+(x_2 - \mu_2)^T[(\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12})^{-1}](x_2 - \mu_2)$$

Now we conduct another *Theorem* to display the $D(x_1,\ x_2)$ in a more compact form.

Theorem 3

Given symmetric matrix equation $A = A^T$, for any vectors u and v we have:

$$u^T A u - 2u^T A v + v^T A v = (v - u)^T A(v - u)$$

Applying *Theorem 3* on $D(x_1,\ x_2)$:

$$D(x_1,\ x_2) = (x_1 - \mu_1)^T\Sigma_{11}^{-1}(x_1 - \mu_1)$$

$$+[(x_2 - \mu_2) - \Sigma_{12}^T\Sigma_{11}^{-1}(x_1 - \mu_1)]^T(\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12})^{-1}[(x_2 - \mu_2) - \Sigma_{12}^T\Sigma_{11}^{-1}(x_1 - \mu_1)]$$

Let's define

$$\boxed{A = \Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}}$$

$$\boxed{b = \mu_2 + \Sigma_{12}^T\Sigma_{11}^{-1}(x_1 - \mu_1)}$$

and

$$D_1(x_1) = (x_1 - \mu_1)^T\Sigma_{11}^{-1}(x_1 - \mu_1)$$
$$D_2(x_1,\ x_2) = (x_2 - b)^T A^{-1}(x_2 - b)$$

Thus we will have

$$D(x_1,\ x_2) = D_1(x_1) + D_2(x_1,\ x_2)$$

Now we can rewrite the joint distribution as

$$p(x_1,\ x_2 \mid \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} e^{\frac{-1}{2}D(x_1,\ x_2)}$$

15

Now according to the *Theorem 3* we get

$$\left|\Sigma\right| = \left|\Sigma_{11}\right|\left|\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}\right|$$

Replacing in the joint probability we get

$$p(x_1,\ x_2 \mid \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma_{11}|^{\frac{1}{2}}|\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}|^{\frac{1}{2}}}e^{\frac{-1}{2}D(x_1,\ x_2)}$$

Separating x_1 and x_2 according to the definitions above, we'll have the following result for each of the random variables (*this proof is for the general mode, the question has asked for the simple 2*2 mode.*)

$$p(x_1,\ x_2 \mid \mu,\ \Sigma) = \mathcal{N}_{x_1}(\mu_1,\ \Sigma_{11})\mathcal{N}_{x_2}(b,\ A)$$

which in this special case, we have

$$X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad M = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} \quad and \quad \Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

Thus, the *separated distribution* of x_1 and x_2 will be:

$$p(x_1,\ x_2 \mid \mu,\ \Sigma) = \mathcal{N}_{x_1}(m_1,\ \sigma_1^2)\mathcal{N}_{x_2}(b,\ A)$$

Finally, the *marginal density* of x_1 is

$$p(x_1) = \int p(x_1,\ x_2)\mathrm{d}x_2 = \frac{1}{2\pi|\sigma_1^2|^{\frac{1}{2}}}e^{\frac{-1}{2}(x_1-m_1)^T\Sigma^{-1}(x_1-m_1)}$$

Thus we get the following *marginal densities* as result.

$$p(x_1) \sim \mathcal{N}_{x_1}(m_1,\ \sigma_1^2)$$

$$p(x_2) \sim \mathcal{N}_{x_2}(m_2,\ \sigma_2^2)$$

For the *conditional density* of $p(x_1 \mid x_2)$ we will divide the *joint probability* to the forsaken *marginal probability*. Replacing b and A from the problem description we get:

$$p(x_1 \mid x_2) = \frac{p(x_1,\ x_2)}{p(x_2)} = \frac{1}{2\pi|\sigma_2^2 - (\rho\sigma_1\sigma_2)^2(\sigma_1)^2|^{\frac{1}{2}}}e^{\frac{-1}{2}(x_2-m_2-(\rho\sigma_1\sigma_2)(\sigma_1)^2)^T\Sigma^{-1}(x_2-m_2-(\rho\sigma_1\sigma_2)(\sigma_1)^2)}$$

Which denotes that

$$p(x_1 \mid x_2) \sim \mathcal{N}_{x_1}(m_1 + \frac{\rho\sigma_1(x_2 - m_2)}{\sigma_2},\ \sigma_1^2(1 - \rho^2))$$

16