A Characterization of the Normal Conditional Distributions

Kazuhiko Matsuno

Contents
1 Introduction
2 Bivariate Normal Distribution
3 Preparation 1
4 Consistent Bivariate Distribution
5 Multivariate Normal Distribution
6 Preparation 2
7 Consistent Multivariate Distribution
8 Extension
9 Concluding Remarks

1 Introduction

It is well known that the conditional distributions derived from a bivariate normal distribution are normal. That is, a bivariate normal distribution yields a pair of normal conditional distributions. The expectations of the normal conditional distributions are called a regression function, which is a linear function of a conditioning variable. The variances of the conditional distributions are constant and called homoscedastic. However, the reverse problem might not be well discussed. That is, if two conditional distributions are first given normal, then a natural question arises; whether the both normal conditional distributions come from a bivariate normal distribution? We simply tend to answer this
reverse question affirmatively based on the relationship between the normal conditional and joint distributions mentioned above. The question does not seem to have gained a proper attention, and the affirmative answer is not given an appropriate proof. The paper takes up this problem.

Conditional distributions are obtained from a joint distribution through a routine way. Given a joint distribution, we derive marginal distributions by summation or integration. Furthermore by division of joint distribution by marginal ones, we get conditional distributions. This is a direct and forward way of obtaining conditional distributions. In this paper we will consider the reverse and backward problem of obtaining a joint distribution and marginal distributions from a pair of conditional distributions. The problem is analyzed under the setting of normality.

Normal distributions are theoretically and in applications most important in Statistics. We already have many interesting properties of the distribution. For an example, a linear combination of normal variates is distributed as normal, and the converse also holds. Another example, as mentioned above, is that conditional distributions from a bivariate normal distribution are also normal. This paper intends to add still another characterization of the normal distributions.

The analysis leads us to a consideration of an integral equation of a particular type, though integral equations are rare in Statistics. Solving the integral equation leads to a consideration of a matrix equation. Furthermore, solving the matrix equation leads to a notion of diagonalization of symmetric or non-symmetric matrices and finally to the notion of Jordan canonical form. It is shown that the solution of the matrix equation turns out to be given by matrix series or an infinite sum of matrices. We consider the condition for the convergence of the matrix series.

Section 2 sets forth the problem within the framework of bivariate distributions and a pair of conditionally normal distributions. In Section 3, we develop necessary mathematical apparatus. In Section 4 we obtain a
condition for that a pair of normal conditional distributions are yielded from a bivariate normal distribution. In Section 5, the problem is specified within a multivariate distribution framework. In Section 6, we consider a matrix equation, notions of matrix diagonalization, and matrix series to define a solution of the matrix equation. Section 7 gives a condition for that a pair of multivariate normal conditional distributions are yielded from a multivariate normal distribution. Section 8 is a final extension towards cases where the key matrices are not necessarily diagonalized. And it is shown that the notion of Jordan canonical form is of much help to extend the results obtained in the previous sections.

2 Bivariate Normal Distribution

Let us consider two random variables $X$ and $Y$, whose joint distribution is described by a joint probability density function $f(x, y)$. The corresponding conditional density functions are derived from $f(x, y)$ and denoted by $f_x(x|y)$ and $f_y(y|x)$. On the contrary, we consider in this paper the situation where a pair of conditional density function are given first, and try to find their common joint distribution. There might not be the common joint distribution. Under what condition does the common joint distribution exist? We take up this problem within a framework where the conditional distributions are both normal.

To fix the idea and notation before starting the analysis, it is helpful to recollect some properties about the normal distribution. If a random variable $X$ is distributed as normal with mean $\mu$ and variance $\sigma^2$, its probability density function is written as

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

When we write this distribution as $N(\mu, \sigma^2)$, it is to be understood that this includes the restrictions on the parameters, $-\infty < \mu < +\infty$ and $0 < \sigma^2 < +\infty$.

If two random variables $X$ and $Y$ are distributed as a bivariate normal distribution, its probability density function is written as
\[ f(x, y) = k \exp \left( -\frac{1}{2} Q(x, y) \right), \]

where

\[
k = (2\pi)^{-1} \left[ \omega_{11} \omega_{22} (1 - \rho^2) \right]^{-1/2},
\]

\[
Q(x, y) = \frac{1}{1 - \rho^2} \left[ \frac{(x - \mu_1)^2}{\omega_{11}} - 2\rho \frac{(x - \mu_1)(y - \mu_2)}{(\omega_{11}\omega_{22})^{1/2}} + \frac{(y - \mu_2)^2}{\omega_{22}} \right].
\]

We have the restrictions on the parameters \(-\infty < \mu_1, \mu_2 < +\infty, 0 < \omega_{11}, \omega_{22} < +\infty\) and \(0 < \omega_{11}\omega_{22} - \omega_{12}^2\), and \(\rho^2 = \omega_{12}^2 / (\omega_{11}\omega_{22}) < 1\). This joint distribution is denoted as \(N(\mu, \Omega)\) including the restrictions on the parameters, or as

\[
N \left( \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{bmatrix} \right), \tag{1}
\]

where the matrix elements are explicit.

When the joint normal distribution (1) is given, the conditional distribution of \(X\) given \(Y = y\) and that of \(Y\) given \(X = x\), respectively, are both normal:\(^{1})

\[
f_x(x|y) = N \left( \mu_1 + \omega_{12} \omega_{22}^{-1} (y - \mu_2), \omega_{11} (1 - \rho^2) \right), \tag{2}
\]

\[
f_y(y|x) = N \left( \mu_2 + \omega_{12} \omega_{11}^{-1} (x - \mu_1), \omega_{22} (1 - \rho^2) \right). \tag{3}
\]

It is noted that we will not deal with the degenerated cases where \(\omega_{11} = 0\), \(\omega_{22} = 0\) and/or \(\rho^2 = 1\).

Letting the regression coefficients of (2) and (3)

\[
\beta = \omega_{12} \omega_{22}^{-1}, \delta = \omega_{12} \omega_{11}^{-1},
\]

---

we have restrictions on the coefficients

\[ \beta \omega_{22} (\equiv \omega_{12}) = \delta \omega_{11}, \]  
\[ 0 \leq \beta \delta (= \rho^2) < 1. \]  

The restrictions (4) and (5) are necessary conditions for that the joint distribution (1) yields the conditional distributions (4) and (5). Our problem is to show if the necessary conditions are sufficient or not.

It is noted that we have, from (5),

\[ \text{sgn}(\beta) = \text{sgn}(\delta). \]

### 2.1 The problem

Now, let us specify a normal conditional distribution of a random variable \( X \) given \( Y = y \) and that of \( Y \) given \( X = x \), respectively, in general terms

\[ f_x(x | y) = N(\alpha + \beta(y - \gamma), \sigma_{11}), \]  
\[ f_y(y | x) = N(\gamma + \delta(x - \alpha), \sigma_{22}). \]

And the questions are asked. Does this pair of conditional distributions come from a certain bivariate normal distribution? What is the condition for the affirmative answer? And what property would this bivariate normal distribution have? In other words we consider a relation of a joint distribution \( N(\mu, \Omega) \) with the conditional distributions (6) and (7). In fact, we have shown above that a necessary condition for \( N(\mu, \Omega) \) to yield normal conditional distributions. Furthermore, we will obtain a sufficient condition for (6) and (7) to be yielded from \( N(\mu, \Omega) \).

The problem is provided for a bivariate distribution of \( X \) and \( Y \) without the normality. Letting a marginal distribution \( g(x) \) of a random variable \( X \) be given, we have

\[ g(x) = \int f(x, y) dy \]
Similarly for a marginal distribution $h(y)$ of a random variable $Y$, we have

$$h(y) = \int f_y(y|w)g(w)dw. \quad (9)$$

Then, substituting (9) into (8) and arranging the result, we get an integral equation for $g(x)$,

$$g(x) = \int f_x(x|y) \left[ \int f_y(y|w)g(w)dw \right] dy$$

$$= \int \left[ \int f_x(x|y)f_y(y|w)dy \right] g(w)dw$$

$$= \int [H(x, w)] g(w)dw, \quad (10)$$

where

$$H(x, w) = \int f_x(x|y)f_y(y|w)dy.$$  

The kernel function $H(x, w)$ is defined, if the conditional density functions $f_x(x|y)$ and $f_y(y|w)$ are given. So, the problem turns out to be solving the equation (10) for $g(x)$. Specifically for the normal distribution cases, the functions $f_x(x|y)$ and $f_y(y|x)$ are specified in (6) and (7). The solution $g(x)$ is called consistent with $f_x(x|y)$ and $f_y(y|x)$.

3 Preparation 1

We here show some properties concerning the normal density functions.

3.1 Properties of the normal density

**Lemma 1** Provided two conditional distributions, concerning three random variables $X$, $Y$ and $W$, 

$$= \int f_x(x|y)h(y)dy. \quad (8)$$
where $0 < a, 0 < c$, we get a conditional distribution $f_x(x|w) = N(bdw, a + b^2c)$.

**Proof.** The density functions of the conditional distributions are

$$f_x(x|y) = k_1 \exp \left\{-\frac{1}{2a} (x - by)^2\right\},$$

$$f_y(y|w) = k_2 \exp \left\{-\frac{1}{2c} (y - dw)^2\right\},$$

where $k_1 = (2\pi a)^{-1/2}$ and $k_2 = (2\pi c)^{-1/2}$. Making the product of them, we have

$$f_x(x|y)f_y(y|w) = k_1k_2 \exp \left\{-\frac{1}{2ac} (cx^2 + daw^2)\right\}$$

$$\times \exp \left\{ \frac{1}{2ac} \left[ \frac{(daw + bcx)^2}{(a + b^2c)} \right] \right\}$$

$$\times \exp \left\{ -\frac{a + b^2c}{2ac} \left[ y - \frac{(daw + bcx)}{(a + b^2c)} \right]^2 \right\}.$$

After integrating this with respect to $y$, we obtain

$$H(x, w) = \int_{-\infty}^{\infty} f_x(x|y)f_y(y|w)dy$$

$$= k_1k_2 \exp \left\{-\frac{1}{2ac} (cx^2 + daw^2)\right\}$$

$$\times \exp \left\{ \frac{1}{2ac} \left[ \frac{(daw + bcx)^2}{(a + b^2c)} \right] \right\}$$

$$\times \int_{-\infty}^{\infty} \exp \left\{ -\frac{a + b^2c}{2ac} \left[ y - \frac{(daw + bcx)}{(a + b^2c)} \right]^2 \right\} dy,$$

where the integral in the last line is $(2\pi ac/(a+b^2c))^{1/2}$. So we get, after rearrangement,

$$H(x, w) = k_3 \exp \left\{-\frac{1}{2(a + b^2c)} (x - bdw)^2\right\},$$

where $k_3 = (2\pi(a+b^2c))^{-1/2}$ and $0 < a+b^2c$. The function $H(x, w)$ is a density.
function of a conditional distribution \( N(bdw, a+b^2c) \) of \( X \) given \( W=w \).

From Lemma 1, we get the following proposition.

**Lemma 2** Conditional distributions \( f_x(x|y) = N(by, a) \) and \( f_y(y|w) = N(0, c) \) together yield a marginal distribution \( N(0, a+b^2c) \) of \( X \).

**Proof.** Putting \( d = 0 \) in Lemma 1, we obtain the result. ■

### 3.2 Integral equation

From Lemma 2, we get the next proposition about the integral equation of our concern.

**Lemma 3** Letting the kernel function \( H(x, w) \) of the integral equation (10) be

\[
H(x, w) = \frac{1}{\sqrt{2\pi a}} \exp\left(-\frac{1}{2a}(x - bw)^2\right) = N(bw, a),
\]

where \( 0 < a \) and \( 0 < 1 - b^2 \), then the function \( g(x) \)

\[
g(x) = \frac{1}{\sqrt{2\pi a/(1-b^2)}} \exp\left(-\frac{1}{2a/(1-b^2)}x^2\right) = N(0, a/(1-b^2)),
\]

is a solution for the integral equation.

**Proof.** Let \( H(x, w) = N(bw, a), G(w : c) = N(0, c) \) and \( G(x : a+b^2c) = N(0, a+b^2c) \). Lemma 2 implies that the conditional distribution \( H(x, w) \) and the marginal distribution \( G(w : c) \) yield the marginal distribution \( G(x : a+b^2c) \). That is, this implies that the equation

\[
G(x : a+b^2c) = \int H(x, w)G(w : c)dw,
\]

holds. If we let \( a+b^2c = c \) or \( c = a/(1-b^2) \), under the assumption \( 0 < 1 - b^2 \), \( G(x : a+b^2c) \) can be rewritten as \( G(x : a/(1-b^2)) \). Then the equation (11) is rewritten as

\[
G(x : a/(1-b^2)) = \int H(x, w)G(w : a/(1-b^2))dw.
\]