Chapter 13

Multivariate normal distributions

The multivariate normal is the most useful, and most studied, of the standard joint distributions. A huge body of statistical theory depends on the properties of families of random variables whose joint distributions are at least approximately multivariate normal. The bivariate case (two variables) is the easiest to understand, because it requires a minimum of notation. Vector notation and matrix algebra becomes necessities when many random variables are involved: for random variables $X_1, \dots, X_n$ write $\mathbf{X}$ for the ***random vector*** $(X_1, \dots, X_n)$, and $\mathbf{x}$ for the generic point $(x_1, \dots, x_n)$ in $\mathbb{R}^n$.

Remark. In general, if $W = (W_{ij})$ is an $m \times n$ matrix whose elements are random variables, the $m \times n$ matrix $\mathbb{E}W$ is defined to have (i,j)th element $\mathbb{E}W_{ij}$. That is, expectations are taken element-wise. If B is an $n \times p$ matrix of constants then WB has (i,j)th element $\sum_{k=1}^{n} W_{ik}B_{kj}$ whose expected value equals $\sum_{k=1}^{n}(\mathbb{E}W_{ik})B_{\ell j}$, the (i,j)th element of the matrix $(\mathbb{E}W)B$. That is, $\mathbb{E}(WB) = (\mathbb{E}W)B$. Similarly, for an $\ell \times m$ matrix of constants A, the expected value of AW equals $A(\mathbb{E}W)$.

For a $1 \times n$ vector of random variables $\mathbf{X} = (X_1, \dots, X_n)$, with expected value $\boldsymbol{\mu} = \mathbb{E}X$, the variance matrix $\text{var}(X)$ is defined to be the $n \times n$ matrix $\mathbb{E}(\mathbf{X} - \boldsymbol{\mu})'(\mathbf{X} - \boldsymbol{\mu})$, whose (i,j)th element equals $\mathbb{E}(X_i - \mu_i)(X_j - \mu_j) = \text{cov}(X_i, X_j)$.

For random vectors $\mathbf{x}$ and $\mathbf{Y}$ with expected values $\boldsymbol{\mu}_X$ and $\boldsymbol{\mu}_Y$, the covariance matrix equals $\mathbb{E}(\mathbf{X} - \boldsymbol{\mu}_X)'(\mathbf{Y} - \boldsymbol{\mu}_Y)$, whose (i,j)th elements equals $\text{cov}(X_i, Y_j)$.

As an exercise you should check that, for an $n \times p$ matrix B of constants, $\text{var}(\mathbf{X}B) = B'\text{var}(\mathbf{X})B$. Other results for variance (and covariance matrices) can be derived similarly.

Be careful when checking these definitions against Wikipedia. I have made my random vectors row vectors; some authors use column vectors.

Definition. Random variables $X_1, X_2, \dots, X_n$ are said to have a jointly continuous distribution with joint density function $f(x_1, x_2, \dots, x_n)$ if, for each subset A of $\mathbb{R}^n$,

$$\begin{aligned}\mathbb{P}\{\mathbf{X} \in A\} &= \iint \dots \int \{(x_1, x_2, \dots x_n) \in A\} f(x_1, x_2, \dots, x_n)\, dx_1\, dx_2 \dots dx_n \\ &= \int \{\mathbf{x} \in A\} f(\mathbf{x})\, d\mathbf{x},\end{aligned}$$

where $\int \dots d\mathbf{x}$ is an abbreviation for the n-fold integral.

For small regions Δ containing a point $\mathbf{x}$,

$$\frac{\mathbb{P}\{\mathbf{X} \in \Delta\}}{vol(\Delta)} \to f(\mathbf{x}) \qquad \text{as } \Delta \text{ shrinks down to } \mathbf{x}.$$

Here $vol(\Delta)$ denotes the n-dimensional volume of Δ.

The density f must be nonnegative and integrate to one over $\mathbb{R}^n$. If the random variables $X_1, \dots, X_n$ are independent, the joint density function is equal to the product of the marginal densities for each X_i, and conversely. The proof is similar to the proof for the bivariate case. For example, if $Z_1, \dots, Z_n$ are independent and each Z_i has a $N(0,1)$ distribution, the joint density is

$$\begin{aligned} f(z_1, \dots, z_n) &= (2\pi)^{-n/2} \exp\left(-\sum\nolimits_{i \le n} z_i^2/2\right) \qquad \text{for all } z_1, \dots, z_n \\ &= (2\pi)^{-n/2} \exp(-\|\mathbf{z}\|^2/2) \qquad \text{for all } \mathbf{z}. \end{aligned}$$

This joint distribution is denoted by $N(\mathbf{0}, I_n)$. It is often referred to as the ***spherical normal distribution***, because of the spherical symmetry of the density. The $N(\mathbf{0}, I_n)$ notation refers to the vector of means and the variance matrix,

$$\mathbb{E}\mathbf{Z} = (\mathbb{E}Z_1, \dots, \mathbb{E}Z_n) = \mathbf{0} \qquad \text{AND} \qquad \text{var}(\mathbf{Z}) = I_n.$$

Remark. More generally, if $X = \boldsymbol{\mu} + \mathbf{Z}A$, where $\boldsymbol{\mu}$ is a constant vector in $\mathbb{R}^n$ and A is a matrix of constants and $\mathbf{Z} = N(\mathbf{0}, I_n)$, then

$$\mathbb{E}\mathbf{X} = \mu \qquad \text{AND} \qquad \text{var}(\mathbf{X}) = V = A'A.$$

If the variance matrix V is non-singular, the n-dimensional analog of the methods in Chapter 11 show that $\mathbf{X}$ has joint density

$$f(\mathbf{x}) = (2\pi)^{-n/2} |\det(V)|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})V^{-1}(\mathbf{x} - \boldsymbol{\mu})'\right)$$

This distribution is denoted by $N(\boldsymbol{\mu}, V)$.

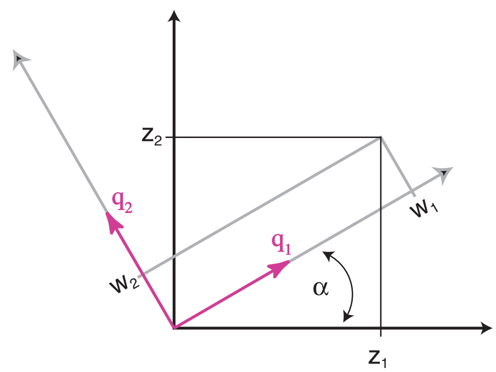
You don't really need to know about the general $N(\boldsymbol{\mu}, V)$ density for this course.

The distance of the random vector $\mathbf{Z}$ from the origin is $\|\mathbf{Z}\| = \sqrt{Z_1^2 + \cdots + Z_n^2}$. From Chapter 11, if $\mathbf{Z} \sim N(\mathbf{0}, I_n)$ you know that $\|\mathbf{Z}\|^2/2$ has a gamma$(n/2)$ distribution. The distribution of $\|\mathbf{Z}\|^2$ is given another special name, because of its great importance in the theory of statistics.

Definition. Let $\mathbf{Z} = (Z_1, Z_2, \dots, Z_n)$ have a spherical normal distribution, $N(\mathbf{0}, I_n)$. The ***chi-square***, χ_n^2, is defined as the distribution of $\|\mathbf{Z}\|^2 = Z_1^2 + \cdots + Z_n^2$.

The methods for finding (bivariate) joint densities for functions of two random variables with jointly continuous distributions extend to multivariate distributions. Admittedly there is a problem with the drawing of pictures in n dimensions, to keep track of the transformations, and one must remember to say "n-dimensional volume" instead of area, but otherwise calculations are not much more complicated than in two dimensions.

The spherical symmetry of the $N(\mathbf{0}, I_n)$ makes some arguments particularly easy. Let me start with the two-dimensional case. Suppose Z_1 and Z_2 have independent $N(0,1)$ distributions, defining a random point $\mathbf{Z} = (Z_1, Z_2)$ in the plane. You could also write $\mathbf{Z}$ as $Z_1\mathbf{e}_1 + Z_2\mathbf{e}_2$, where $\mathbf{e}_1 = (1,0)$ and $\mathbf{e}_2 = (0,1)$. Rotate the coordinate axes through an angle α, writing $\mathbf{W} = (W_1, W_2)$ for the coordinates of the random point in the new coordinate system.



The new axes are defined by the unit vectors

$$\mathbf{q}_1 = (\cos\alpha, \sin\alpha) \qquad \text{AND} \qquad \mathbf{q}_2 = (-\sin\alpha, \cos\alpha).$$

Remark. Note that $\mathbf{q}_1$ and $\mathbf{q}_2$ are orthogonal because $\mathbf{q}_1 \cdot \mathbf{q}_2 = 0$.

The representation $\mathbf{Z} = (Z_1, Z_2) = W_1\mathbf{q}_1 + W_2\mathbf{q}_2$ gives

$$W_1 = \mathbf{Z} \cdot \mathbf{q}_1 = Z_1 \cos\alpha + Z_2 \sin\alpha$$
$$W_2 = \mathbf{Z} \cdot \mathbf{q}_2 = -Z_1 \sin\alpha + Z_2 \cos\alpha.$$

That is, W_1 and W_2 are both linear functions of Z_1 and Z_2. The random variables $\mathbf{W} = (W_1, W_2)$ have a multivariate normal distribution with $\mathbb{E}\mathbf{W} = \mathbf{0}$ and

$$\begin{aligned}\text{var}(W_1) &= \cos^2\alpha + \sin^2\alpha = 1\\ \text{var}(W_2) &= \sin^2\alpha + \cos^2\alpha = 1\\ \text{cov}(W_1, W_2) &= (\cos\alpha)(-\sin\alpha) + (\sin\alpha)(\cos\alpha) = 0.\end{aligned}$$

More succinctly, $\text{var}(\mathbf{W}) = I_2$, a property that you could check more cleanly using the representation $\mathbf{W} = \mathbf{Z}Q'$, where Q is the orthogonal matrix with rows $\mathbf{q}_1$ and $\mathbf{q}_2$. In fact, the random variables W_1 and W_2 are independent and each is distributed $N(0,1)$. I won't give all the details for the two-dimensional case because the argument in higher dimensions also works for $\mathbb{R}^2$.

<1> **Example.** Suppose $\mathbf{Z} \sim N(\mathbf{0}, I_n)$. Let $\mathbf{q}_1, \dots, \mathbf{q}_n$ be a new orthonormal basis for $\mathbb{R}^n$, and let $\mathbf{Z} = W_1\mathbf{q}_1 + \cdots + W_n\mathbf{q}_n$ be the representation for $\mathbf{Z}$ in the new basis. Then the $W_1, \dots, W_n$ are also independent $N(0,1)$ distributed random variables.

To prove results involving the spherical normal it is often merely a matter of transforming to an appropriate orthonormal basis. This technique greatly simplifies the study of statistical problems based on multivariate normal models.

<2> **Example.** Suppose $Z_1, Z_2, \dots, Z_n$ are independent, each distributed $N(0,1)$. Define $\bar{Z} = (Z_1 + \cdots + Z_n)/n$ and $T = \sum_{i \le n}(Z_i - \bar{Z})^2$. Show that $\bar{Z}$ has a $N(0, 1/n)$ distribution independently of T, which has a χ^2_{n-1} distribution.

Statistical problems often deal with independent random variables $Y_1, \dots, Y_n$ each distributed $N(\mu, \sigma^2)$, where μ and σ^2 are unknown parameters that need to be estimated. If we define $Z_i = (Y_i - \mu)/\sigma$ then the Z_i are as in the previous Example. Moreover,

$$\begin{aligned} \bar{Y} &= \frac{1}{n}\sum\nolimits_{i \le n} Y_i = \mu + \sigma\bar{Z} \sim N(\mu, \sigma^2/n) \\ \sum\nolimits_{i \le n}(Y_i - \bar{Y})^2/\sigma^2 &= \sum\nolimits_{i \le n}(Z_i - \bar{Z})^2 \sim \chi^2_{n-1}, \end{aligned}$$

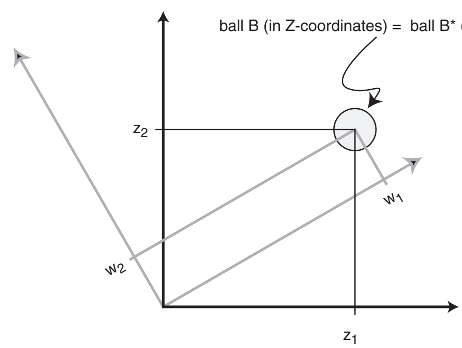
from which it follows that $\bar{Y}$ and $\widehat{\sigma}^2 := \sum_{i \le n}(Y_i - \bar{Y})^2/(n-1)$ are independent.

Remark. It is traditional to use $\bar{Y}$ to estimate μ and $\widehat{\sigma}^2$ to estimate σ^2. The random variable $\sqrt{n}(\bar{Y} - \mu)/\widehat{\sigma}$ has the same distribution as $U/\sqrt{V/(n-1)}$, where $U \sim N(0,1)$ independently of $V \sim \chi^2_{n-1}$. By definition, such a ratio is said to have a ***t distribution on*** $n-1$ ***degrees of freedom***.

<3> **Example.** Distribution of least squares estimators for regression.

Examples for Chapter 13

Example 1
We have $\mathbf{Z} \sim N(\mathbf{0}, I_n)$ and $\mathbf{q}_1, \dots, \mathbf{q}_n$ a new orthonormal basis for $\mathbb{R}^n$. In the new coordinate system, $\mathbf{Z} = W_1\mathbf{q}_1 + \cdots + W_n\mathbf{q}_n$ We need to show that the $W_1, \dots, W_n$ are also independent $N(0,1)$ distributed random variables.



The picture shows only two of the n coordinates; the other $n-2$ coordinates are sticking out of the page. I have placed the pictures for the **w**- and **z**-spaces on top of each other, so that you can see how the balls B and B^* line up.

For a small ball B centered at $\mathbf{z}$,

$$\mathbb{P}\{\mathbf{Z} \in B\} \approx f(\mathbf{z})(\text{volume of } B) \qquad \text{where } f(\mathbf{z}) = (2\pi)^{-n/2}\exp(-\|\mathbf{z}\|^2/2).$$

The corresponding region for $\mathbf{W}$ is B^*, a ball of the same radius, but centered at the point $\mathbf{w} = (w_1, \dots, w_n)$ for which $w_1\mathbf{q}_1 + \cdots + w_n\mathbf{q}_n = \mathbf{z}$. Thus

$$\mathbb{P}\{\mathbf{W} \in B^*\} = \mathbb{P}\{\mathbf{Z} \in B\} \approx (2\pi)^{-n/2}\exp(-\tfrac{1}{2}\|\mathbf{x}\|^2)(\text{volume of } B).$$

From the equalities

$$\|\mathbf{w}\| = \|\mathbf{z}\| \qquad \text{AND} \qquad \text{volume of } B = \text{volume of } B^*,$$

we get

$$\mathbb{P}\{\mathbf{W} \in B^*\} \approx (2\pi)^{-n/2}\exp(-\tfrac{1}{2}\|\mathbf{w}\|^2)(\text{volume of } B^*).$$

That is, $\mathbf{W}$ has the asserted $N(\mathbf{0}, I_n)$ density. $\square$

Example 2
Suppose $Z_1, Z_2, \dots, Z_n$ are independent, each distributed $N(0,1)$. Define

$$\bar{Z} = \frac{Z_1 + \cdots + Z_n}{n} \qquad \text{and} \qquad T = \sum\nolimits_{i \le n} (Z_i - \bar{Z})^2$$

Show that $\bar{Z}$ has a $N(0, 1/n)$ distribution independently of T, which has a χ^2_{n-1} distribution.

Choose the new orthonormal basis with $\mathbf{q}_1 = (1, 1, \dots, 1)/\sqrt{n}$. Choose $\mathbf{q}_2, \dots, \mathbf{q}_n$ however you like, provided they are orthogonal unit vectors, all orthogonal to $\mathbf{q}_1$. In the new coordinate system,

$$\mathbf{Z} = W_1\mathbf{q}_1 + \cdots + W_n\mathbf{q}_n \qquad \text{where } W_i = \mathbf{Z} \cdot \mathbf{q}_i \text{ for each } i.$$

In particular,

$$W_1 = \mathbf{Z} \cdot \mathbf{q}_1 = \frac{Z_1 + \cdots + Z_n}{\sqrt{n}} = \sqrt{n}\bar{Z}$$

From Example <1> you know that W_1 has a $N(0,1)$ distribution. It follows that $\bar{Z}$ has a $N(0, 1/n)$ distribution.

The random variable T equals the squared length of the vector

$$(Z_1 - \bar{Z}, \dots, Z_n - \bar{Z}) = \mathbf{Z} - \bar{Z}(\sqrt{n}\mathbf{q}_1) = \mathbf{Z} - W_1\mathbf{q}_1 = W_2\mathbf{q}_2 + \cdots + W_n\mathbf{q}_n.$$

That is,

$$T = \|W_2\mathbf{q}_2 + \cdots + W_n\mathbf{q}_n\|^2 = W_2^2 + \cdots + W_n^2,$$

a sum of squares of $n-1$ independent $N(0,1)$ random variables, which has a χ^2_{n-1}-distribution.

Finally, notice that $\bar{Z}$ is a function of W_1, whereas T is a function of the independent random variables $W_2, \dots, W_n$. The independence of $\bar{Z}$ and T follows. $\square$

Example 3
Suppose $Y_1, \dots Y_n$ are independent random variables, with $Y_i \sim N(\mu_i, \sigma^2)$ for an unknown σ^2. Suppose also that $\mu_i = \alpha + \beta x_i$, for unknown parameters α and β and observed constants $x_1, \dots, x_n$ with average $\bar{x} = \sum_{i \le n} x_i/n$.

The method of least squares estimates the parameters α and β by the values $\widehat{a}$ and $\widehat{b}$ that minimize

$$S^2(a,b) = \sum\nolimits_{i \le n} \left(Y_i - a - bx_i\right)^2$$

over all (a,b) in $\mathbb{R}^2$. One then estimates σ^2 by the value $\widehat{\sigma}^2 = S^2(\widehat{a}, \widehat{b})/(n-2)$.

In what follows I will assume that $T := \sum_{i=1}^n (x_i - \bar{x})^2 > 0$. (If T were zero then all the x_i would be equal, which would make $\mathbf{x} = \bar{x}\mathbf{1}$.)

Define $\mathbf{Y} = (Y_1, \dots, Y_n)$ and $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{1} = (1, 1, \dots, 1)$. Then

$$\mathbb{E}\mathbf{Y} = \boldsymbol{\mu} = \alpha\mathbf{1} + \beta\mathbf{x} \qquad \text{AND} \qquad \mathbf{Y} = \boldsymbol{\mu} + \sigma\mathbf{Z} \qquad \text{where } \mathbf{Z} \sim N(\mathbf{0}, I_n)$$

and

$$S^2(a, b) = \|\mathbf{Y} - a\mathbf{1} - b\mathbf{x}\|^2 .$$

Create a new orthonormal basis for $\mathbb{R}^n$ by taking

$$\mathbf{q}_1 = (1, 1, \dots, 1)/\sqrt{n} \qquad \text{AND} \qquad \mathbf{q}_2 = \frac{\mathbf{x} - \bar{x}\mathbf{1}}{\|\mathbf{x} - \bar{x}\mathbf{1}\|}$$

Choose $\mathbf{q}_3, \dots, \mathbf{q}_n$ however you like, provided they are orthogonal unit vectors, all orthogonal to $\mathbf{q}_1$.

Remark. You should check that $\mathbf{q}_1 \cdot \mathbf{q}_2 = 0$ and $\|\mathbf{q}_1\| = \|\mathbf{q}_2\| = 1$. Also note that $\|\mathbf{x} - \bar{x}\mathbf{1}\| = \sqrt{T}$.

The vectors $\mathbf{1}, \mathbf{x}$ and $\mathbf{q}_1, \mathbf{q}_2$ span the same two-dimensional subspace of $\mathbb{R}^2$. That is, any vector that can be written as a linear combination of $\mathbf{1}$ and $\mathbf{x}$ can also be written as a linear combination of $\mathbf{q}_1$ and $\mathbf{q}_2$; and any vector that can be written as a linear combination of $\mathbf{q}_1$ and $\mathbf{q}_2$ can also be written as a linear combination of $\mathbf{1}$ and $\mathbf{x}$. Put another way, for each pair a, b there is a unique pair c, d for which $a\mathbf{1} + b\mathbf{x} = c\mathbf{q}_1 + d\mathbf{q}_2$.

Remark. In matrix form, $(a, b)X = (c, d)Q$, where X is the 2×2 matrix with rows $\mathbf{1}$ and $\mathbf{x}$, and Q is the 2×2 orthogonal matrix with rows $\mathbf{q}_1$ and $\mathbf{q}_2$. The two matrices are related by $X = RQ$ and $Q = R^{-1}X$ where

$$R = \begin{pmatrix} \sqrt{n} & 0 \\ \sqrt{n}\bar{x} & \sqrt{T} \end{pmatrix} \qquad \text{AND} \qquad R^{-1} = \begin{pmatrix} 1/\sqrt{n} & 0 \\ -\bar{x}/\sqrt{T} & 1/\sqrt{T} \end{pmatrix}.$$

Thus $(a, b)X = (c, d)Q$ if and only if $(a, b)R = (c, d)$ if and only if $(a, b) = (c, d)R^{-1}$. That is,

$$\begin{aligned} c &= \sqrt{n}(a + b\bar{x}), \quad d = \sqrt{T}\, b \\ a &= c/\sqrt{n} - d\bar{x}/\sqrt{T}, \quad b = d/\sqrt{T}. \end{aligned}$$

The calculations for transforming between coordinate systems are easier if you work with matrix notation.

The least squares problem

Write all the vectors in the new basis:

$$\begin{aligned} \widehat{a}\mathbf{1} + \widehat{b}\mathbf{x} &= (\widehat{a} + \widehat{b}\bar{x})\mathbf{1} + \widehat{b}(\mathbf{x} - \bar{x}\mathbf{1}) \\ &= \widehat{c}\mathbf{q}_1 + \widehat{d}\mathbf{q}_2 \qquad \text{where } \widehat{c} = (\widehat{a} + \widehat{b}\bar{x})\sqrt{n} \text{ and } \widehat{d} = \widehat{b}\sqrt{T}, \end{aligned}$$

and

$$\mathbf{Y} = \sum\nolimits_{i\le n} g_i \mathbf{q}_i \qquad \text{where } g_i := \mathbf{Y}\cdot\mathbf{q}_i.$$

Remark. By direct calculation, $g_1 = \mathbf{Y}\cdot\mathbf{1}/\sqrt{n} = \bar{Y}\sqrt{n}$, where $\bar{Y} = \sum_{i\le n} Y_i/n$, and $g_2 = \mathbf{Y}\cdot(\mathbf{x}-\bar{x}\mathbf{1})/\sqrt{T} = \sum_{i\le n} Y_i(x_i-\bar{x})/\sqrt{\sum_{i\le n}(x_i-\bar{x})^2}$.

The quantities $\widehat{c}$ and $\widehat{d}$ minimize, over all $(c,d)\in\mathbb{R}^2$,

$$\begin{aligned}\|\mathbf{Y} - c\mathbf{q}_1 - d\mathbf{q}_2\|^2 &= \Big\|(g_1-c)\mathbf{q}_1 + (g_2-d)\mathbf{q}_2 + \sum\nolimits_{i\ge 3} g_i\mathbf{q}_i\Big\|^2 \\ &= (g_1-c)^2 + (g_2-d)^2 + \sum\nolimits_{i=3}^n g_i^2\end{aligned}$$

Clearly the solution is $\widehat{c} = g_1$ and $\widehat{d} = g_2$. That is,

$$\begin{aligned}\widehat{b} &= \widehat{d}/\sqrt{T} = \sum\nolimits_{i\le n} Y_i(x_i-\bar{x})\Big/\sum\nolimits_{i\le n}(x_i-\bar{x})^2 \\ \widehat{a} &= \widehat{c}/\sqrt{n} - \widehat{d}\bar{x}/\sqrt{T} = \bar{Y} - \widehat{b}\bar{x}\end{aligned}$$

The least squares estimators

By assumption $\mathbf{Y} = \boldsymbol{\mu} + \sigma\mathbf{Z}$ where $\mathbf{Z}\sim N(\mathbf{0}, I_n)$. In the new coordinate system,

$$\mathbf{Z} = W_1\mathbf{q}_1 + W_2\mathbf{q}_2 + \cdots + W_n\mathbf{q}_n \qquad \text{with } \mathbf{W}\sim N(\mathbf{0}, I_n)$$

so that

$$\begin{aligned}\mathbf{Y} &= \boldsymbol{\mu} + \sigma\sum\nolimits_{i=1}^n W_i\mathbf{q}_i \\ &= (\gamma+\sigma W_1)\mathbf{q}_1 + (\delta+\sigma W_2)\mathbf{q}_2 \qquad \text{where } \gamma := (\alpha+\beta\bar{x})\sqrt{n} \text{ and } \delta := \beta\sqrt{T}.\end{aligned}$$

The representation for $\boldsymbol{\mu}$ comes from

$$\mu = \alpha\mathbf{1} + \beta\mathbf{x} = (\alpha+\beta\bar{x})\mathbf{1} + \beta(\mathbf{x}-\bar{x}\mathbf{1}) = \gamma\mathbf{q}_1 + \delta\mathbf{q}_2.$$

Dot both sides of the last equation for $\mathbf{Y}$ with $\mathbf{q}_i$ to get

$$g_i = \mathbf{Y}\cdot\mathbf{q}_i = \begin{cases}\gamma+\sigma W_1 & \text{for } i=1\\ \delta+\sigma W_2 & \text{for } i=2\\ \sigma W_i & \text{for } 3\le i\le n.\end{cases}$$

Thus

$$\begin{aligned}\widehat{c} &= \gamma+\sigma W_1 \sim N(\gamma,\sigma^2)\\ \widehat{d} &= \delta+\sigma W_2 \sim N(\delta,\sigma^2)\\ (n-2)\widehat{\sigma}^2/\sigma^2 &= \Big\|\mathbf{Y} - \widehat{c}\mathbf{q}_1 - \widehat{d}\mathbf{q}_2\Big\|^2/\sigma^2 = \sum\nolimits_{i=3}^n W_i^2 \sim \chi^2_{n-2}.\end{aligned}$$

Moreover, the independence of the W_i's implies that $\widehat{c}$, $\widehat{d}$, and $\widehat{\sigma}^2$ are independent random variables.

More succinctly,

$$(\widehat{c}, \widehat{d}) \sim N\left((\gamma, \delta), \sigma^2 I_i\right),$$

so that

$$(\widehat{a}, \widehat{b}) = (\widehat{c}, \widehat{d})R^{-1} \sim N\left((\alpha, \beta), \sigma^2 (R^{-1})' R^{-1}\right).$$

If you look in a regression textbook you might see the variance matrix rewritten as $\sigma^2 (XX')^{-1}$. □

Remark. All the algebra, including the calculation of matrix inverses and a possible choice for $\mathbf{q}_1, \ldots, \mathbf{q}_n$ is carried out automatically in a statistical package such as R. There is not much point in memorizing the solutions these days.