Multivariate Normal Distribution

Kert Viele

STA601 Spring 2008

Definition

A multivariate vector $X = (X_1, \ldots, X_p)^T$ has a multivariate normal distribution if any of the following equivalent conditions hold

- For any p-dimensional vector a, the linear combination $a^T X$ has a univariate normal distribution
- There exists independent standard normal random variables $Z = (Z_1, \ldots, Z_M)^T$, a p-dimensional vector μ, and a $p \times M$ matrix A such that $X = Az + \mu$
- There exists a vector μ and positive semi-definite matrix Σ such that X has a characteristic function

$$\phi_X(u) = \exp\left\{i\mu^T u - (1/2)u^T \Sigma u\right\}$$

When Σ is nonsingular, the density of X is

$$f_X(x) = (1/2\pi)^{p/2} |\Sigma|^{-1/2} \exp\left\{-(0.5)(x-\mu)^T \Sigma^{-1}(x-\mu)\right\}$$

Notes on the definition

- The first definition, that every linear combination of X is normally distribution, is a good characterization of the distribution. Note having every component of X be normal is NOT sufficient for the vector X to be multivariate normal.
- The second definition is constructive and thus is quite useful for determining properties of the multivariate normal. It says a multivariate normal arises as a linear transformation of iid normal random variables.

A Lemma on Covariance Matrices

Let $W = (W_1, \ldots, W_N)^T$ have covariance matrix Σ (meaning the i, j element of Σ is $Cov(W_i, W_j)$) and A be a $M \times N$ matrix. Then $Cov(y = Aw) = A\Sigma A^T$.

To prove this, we need to show the i, j element of $A\Sigma A^T$ is $Cov(Y_i, Y_j)$. Note $Y_i = a_{i1}W_1 + \ldots + a_{iN}W_N$ and thus

$$Cov(Y_i, Y_j) = Cov(a_{i1}W_1 + \ldots + a_{iN}W_N, a_{j1}W_1 + \ldots a_{jN}W_n)$$

$$= \sum_{k,l} a_{ik}a_{jl}Cov(W_k, W_l) = \sum_{k,l} a_{ik}a_{jl}\sigma_{kl}$$

Going through the tedium of multiplying out $A\Sigma A^T$, we arrive at the same answer

Expectation and Variance

Lemma : Let X have a multivariate normal distribution, and let A be the matrix in the constructive definition. Then $E[x] = \mu$ and $Cov(x) = AA^T$.

According to the constructive definition, $x = Az + \mu$, and thus $E[x] = E[Az + \mu] = E[Az] + \mu = \mu$ and $Cov(Az + \mu) = Cov(Az) = ACov(z)A^T = AA^T$

Typically we refer to $X \sim N_p(\mu, \Sigma)$, or "X has a multivariate normal distribution with mean vector μ and covariance matrix Σ".

Subsets of X

Lemma : Let $X \sim N_p(\mu, \Sigma)$ and let $X' = (X_{i_1}, \ldots, X_{i_{p'}})$ be a subset of the components of X. Then X' is also multivariate normal, with mean vector μ' being the respective components of μ and covariance matrix Σ' with elements corresponding to the respective rows and columns of Σ.

First note that X' may be defined by taking Bx, where B is the p' by p matrix with all elements 0 except each row j has a 1 in the i_j column. The vector X' is multivariate normal because every linear combination of the element of X' is a linear combination of the elements of X, thus the first definition indicates X' is multivariate normal. The mean vector and covariance matrix may be found by taking $B\mu$ and $B\Sigma B^T$ and noting their particular forms.

Singular Covariance Matrices

Let $X \sim N_p(\mu, \Sigma)$. Note that Σ may be singular, indicating a linear dependency among the components of X. Whenever this is true, there exists a subset of the components $X' \sim N_{p'}(\mu', \Sigma')$ such that Σ' is nonsingular and the remaining component of X (those not in X') are linearly determined by X'.

Thus, in what follows, we will assume a nonsingular covariance matrix Σ

Example of Singular Covariance Matrix

Let Z_1, Z_2, Z_3 be independent standard normal variables. Define $X_1 = Z_1 - \bar{Z}$, $X_2 = Z_2 - \bar{Z}$ and $X_3 = Z_3 - \bar{Z}$. This corresponds to

$$X = \begin{bmatrix} 2/3 & -1/3 & -1/3 \\ -1/3 & 2/3 & -1/3 \\ -1/3 & -1/3 & 2/3 \end{bmatrix} \begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \end{pmatrix} + 0$$

In this case, $\Sigma = AA^T$ is singular because of the linear dependency $X_1 + X_2 + X_3 = 0$. However, taking (X_1, X_2) as a vector results in a nonsingular covariance matrix, with $X_3 = -(X_1 + X_2)$.

Finding Probabilities

Let $X \sim N_p(\mu, \Sigma)$ with nonsingular covariance matrix Σ. Recall the third definition states the density of X is

$$f_X(x) = \left(\frac{1}{2\pi}\right)^{\frac{p}{2}} \frac{1}{|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}$$

Thus, we would integrate this density over any set to find the probability of the set. However, like the univariate normal, this cannot be done analytically for most sets and thus requires numerical techniques.

In what follows, we at least want to describe sets analogous to "the middle 50%" or "middle 95%" of the data.

Contours of iid Standard Normals

1. First, let z, A, and μ be the appropriate vectors and matrices from the constructive definition. Note the z components are iid standard normals, and thus $Z_1^2 + \ldots + Z_M^2 \sim \chi^2_M$.
2. Note $Z_1^2 + \ldots + Z_M^2 < c$ is the interior of a circle. Thus, if we want a set with the middle 95% of values, we need to pick c to be the 95^{th} percentile of a χ^2_M. Here $\sqrt{c}$ is the radius of the circle.
3. Note that because the joint density of the z values is higher everywhere inside this circle than outside this circle, this circle is the smallest set that contains 95% of the values.

Contours of Independent Normals

We know by the constructive definition that $x = Az + \mu$. The covariance matrix of X is $\Sigma = AA^T$. This can be decomposed as $\Sigma = D\Lambda D^T$ for an orthonormal D and diagonal Λ. Thus, $AA^T = D\Lambda^{1/2}\Lambda^{1/2}D^T$ and $A = D\Lambda^{1/2}$.

Thus, $x = D\Lambda^{1/2}z + \mu$.

We have discussed the contours of z, so let's move on to the contours of $w = \Lambda^{1/2}z$. Note $W_i = \sqrt{\lambda_i}\, Z_i$. This results in

$$\frac{W_1^2}{\lambda_1} + \ldots + \frac{W_M^2}{\lambda_M} = Z_1^2 + \ldots + Z_M^2 = c$$

The first part of this equation, involving $W_1, \ldots, W_M$, is the equation of an ellipse.

Rotations

Let's start with $w = \Lambda^{1/2} z$. We are about to multiply by an orthonormal matrix D. Note that multiplying by an orthonormal matrix amounts to rotating points.
In two dimensions, note a point $y = (y_1, y_2)^T$ can be expressed in polar coordinates as $y = (r\cos\theta, r\sin\theta)^T$. Suppose we want to rotate the point so that $Dy = [r\cos(\theta+\phi), r\sin(\theta+\phi)]^T$.
Since $\cos(\theta+\phi) = \cos(\theta)\cos(\phi) - \sin(\theta)\sin(\phi)$ and $\sin(\theta+\phi) = \cos(\theta)\sin(\phi) - \sin(\theta)\cos(\phi)$, This results in

$$D = \begin{bmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{bmatrix}$$

Contours of Mean Zero Multivariate Normals

Since going from $w = \Lambda^{1/2} z$ to $x = D\Lambda^{1/2} z$ only involves a rotation, to find a 95% region we have to rotate the ellipse we found for independent normals.

Fortunately, the principle axes of this rotated ellipse are the columns of D. Recall we originally acquired D through the decomposition $\Sigma = D\Lambda D^T$, thus D is the eigenvectors of Σ.