conditional distribution of multi-variate normal variable

Canonical name	ConditionalDistributionOfMultivariateNormalVariable
Date of creation	2013-03-22 18:39:09
Last modified on	2013-03-22 18:39:09
Owner	stevecheng (10074)
Last modified by	stevecheng (10074)
Numerical id	5
Author	stevecheng (10074)
Entry type	Theorem
Classification	msc 62E15
Classification	msc 60E05

Theorem. *Let X be a random variable, taking values in $\mathbb{R}^n$, normally distributed with a non-singular covariance matrix Σ and a mean of zero.*

Suppose Y is defined by $Y = B^* X$ for some linear transformation $B \colon \mathbb{R}^k \to \mathbb{R}^n$ of maximum rank. (* to denotes the transpose operator.)

Then the distribution of X conditioned on Y is multi-variate normal, with conditional means and covariances of:

$$\mathbb{E}[X \mid Y] = \Sigma B (B^* \Sigma B)^{-1} Y \,, \quad \operatorname{Var}[X \mid Y] = \Sigma - \Sigma B (B^* \Sigma B)^{-1} (\Sigma B)^* \,.$$

If $k = 1$, so that B is simply a vector in $\mathbb{R}^n$, these formulas reduce to:

$$\mathbb{E}[X \mid Y] = \frac{\Sigma B Y}{\operatorname{Var}[Y]} \,, \quad \operatorname{Var}[X \mid Y] = \Sigma - \frac{\Sigma B B^* \Sigma}{\operatorname{Var}[Y]} \,.$$

If X does not have zero mean, then the formula for $\mathbb{E}[X \mid Y]$ is modified by adding $\mathbb{E}[X]$ and replacing Y by $Y - \mathbb{E}[Y]$, and the formula for $\operatorname{Var}[X \mid Y]$ *is unchanged.*

Proof. We split up X into two stochastically independent parts, the first part containing exactly the information embodied in Y. Then the conditional distribution of X given Y is simply the unconditional distribution of the second part that is independent of Y.

To this end, we first change variables to express everything in terms of a *standard* multi-variate normal Z. Let $A \colon \mathbb{R}^n \to \mathbb{R}^n$ be a "square root" factorization of the covariance matrix Σ, so that:

$$AA^* = \Sigma \,, \quad Z = A^{-1} X \,, \quad X = AZ \,, \quad Y = B^* A Z \,.$$

We let $H \colon \mathbb{R}^n \to \mathbb{R}^n$ be the orthogonal projection onto the range of $A^* B : \mathbb{R}^k \to \mathbb{R}^n$, and decompose Z into orthogonal components:

$$Z = HZ + (I - H) Z \,.$$

It is intuitively obvious that orthogonality of the two random normal vectors implies their stochastic independence. To show this formally, observe that the Gaussian density function for Z factors into a product:

$$(2\pi)^{-n/2} \exp\bigl(-\tfrac{1}{2} \|z\|^2\bigr) = (2\pi)^{-n/2} \exp\bigl(-\tfrac{1}{2} \|Hz\|^2\bigr) \exp\bigl(-\tfrac{1}{2} \|(I - H) z\|^2\bigr) \,.$$

We can construct an orthonormal system of coordinates on $\mathbb{R}^n$ under which the components for Hz are completely disjoint from those components of

$(I - H)z$. On the other hand, the densities for Z, HZ, and $(I - H)Z$ remain invariant even after changing coordinates, because they are radially symmetric. Hence the variables HZ and $(I - H)Z$ are separable in their joint density and they are independent.

HZ embodies the information in the linear combination $Y = B^*AZ$. For we have the identity:

$$Y = \big(B^*A\big)Z = \big(B^*A\big)\big(HZ + (I - H)Z\big) = \big(B^*A\big)HZ + 0\,.$$

The last term is null because $(I - H)Z$ is orthogonal to the range of A^*B by definition. (Equivalently, $(I - H)Z$ lies in the kernel of $(A^*B)^* = B^*A$.) Thus Y can always be recovered by a linear transformation on HZ.

Conversely, Y completely determines HZ, from the analytical expression for H that we now give. In general, the orthogonal projection onto the range of an injective transformation T is $T(T^*T)^{-1}T^*$. Applying this to $T = A^*B$, we have

$$\begin{aligned} H &= A^*B\big(B^*AA^*B\big)^{-1}B^*A \\ &= A^*B(B^*\Sigma B)^{-1}B^*A\,. \end{aligned}$$

We see that $HZ = A^*B(B^*\Sigma B)^{-1}Y$.

We have proved that conditioning on Y and HZ are equivalent, and so:

$$\mathbb{E}[Z \mid Y] = \mathbb{E}[Z \mid HZ] = \mathbb{E}[HZ + (I - H)Z \mid HZ] = HZ + 0\,,$$

and

$$\begin{aligned} \operatorname{Var}[Z \mid Y] = \operatorname{Var}[Z \mid HZ] &= \operatorname{Var}[HZ + (I - H)Z \mid HZ] \\ &= 0 + \operatorname{Var}[(I - H)Z] \\ &= \mathbb{E}\big[(I - H)ZZ^*(I - H)^*\big] \\ &= (I - H)(I - H)^* \\ &= I - H - H^* + HH^* = I - H\,, \end{aligned}$$

using the defining property $H^2 = H = H^*$ of orthogonal projections.

Now we express the result in terms of X, and remove the dependence on the transformation A (which is not uniquely defined from the covariance matrix):

$$\mathbb{E}[X \mid Y] = A\,\mathbb{E}[Z \mid Y] = AHZ = \Sigma B(B^*\Sigma B)^{-1}Y$$

and

$$\mathrm{Var}[X \mid Y] = A\ \mathrm{Var}[Z \mid Y]\, A^* = AA^* - AHA^* = \Sigma - \Sigma B {(B^* \Sigma B)}^{-1} B^* \Sigma\,.$$

Of course, the conditional distribution of X given Y is the same as that of $(I - H)Z$, which is multi-variate normal.

The formula in the statement of this theorem, for the single-dimensional case, follows from substituting in $\mathrm{Var}[Y] = \mathrm{Var}[B^* X] = B^* \Sigma B$. The formula for when X does not have zero mean follows from applying the base case to the shifted variable $X - \mathbb{E}[X]$. $\square$