Jointly Gaussian Random Variables

Definition

Let $X_1, X_2, \ldots, X_d$ be real valued random variables defined on the same sample space. They are called *jointly Gaussian* if their joint characteristic function is given by

$$\Phi_{\underline{X}}(\underline{u}) = \exp(i\underline{u}^T\underline{m} - \frac{1}{2}\underline{u}^T C\underline{u}) \ . \tag{1}$$

where C is a real, symmetric, nonnegative definite matrix, and $\underline{m} = [m_1, \ldots, m_d]^T \in \mathbf{R}^d$.

If C is positive definite, then one can show that the real valued random variables $X_1, X_2, \ldots, X_d$ are jointly Gaussian iff they have a joint density of the form

$$f_{\underline{X}}(\underline{x}) = \frac{1}{\sqrt{(2\pi)^d \det C}} \exp(-\frac{1}{2}((\underline{x} - \underline{m})^T C^{-1}(\underline{x} - \underline{m}))) \ .$$

Proof It is a simple calculation that the characteristic function associated to the density above is of the form in Eqn. (1). The converse follows from the uniqueness of Fourier inversion. □

However, when C is singular the jointly Gaussian random variables $X_1, X_2, \ldots, X_d$ will not admit a joint density, because the entire joint distribution will be concentrated on the subspace orthogonal to the null space of C.

It is also important to realize that though each of the random variables in a family of jointly Gaussian random variables is necessarily Gaussian, it is possible for random variables to be defined on the sample space, to be individually Gaussian, but to not be jointly Gaussian. For example, consider X and Y jointly distributed with a density that is of the form

$$f_{XY}(x,y) = \frac{1}{2\pi}\exp(-\frac{x^2+y^2}{2}) + \alpha(x-1,y-1) - \alpha(x+1,y-1) + \alpha(x+1,y+1) - \alpha(x-1,y+1)$$

where $\alpha(x, y)$ is a nonnegative function zero outside $\{(x,y) :\mid x \mid, \mid y \mid \leq 1/2$ and $\mid \alpha(x,y) \mid \leq 0.001$ for all (x, y). You can check that $f_{XY}(\cdot, \cdot)$ is a joint density. Then X and Y are each $N(0, 1)$ random variables. However they are not jointly Gaussian.

Characterization via linear combinations

Jointly Gaussian random variables can be characterized by the property that every scalar linear combination of such variables is Gaussian.

Theorem 1 Real valued random variables $X_1, X_2, \ldots, X_d$ are jointly Gaussian iff for all $\underline{a} \in \mathbf{R}^d$, the real r.v. $\sum_i a_i X_i$ is Gaussian.

Proof If $X_1, X_2, \ldots, X_d$ are jointly Gaussian, and $X = \sum_i a_i X_i$, then

$$\begin{aligned} \Phi_X(u) &= E[\exp(iu\sum_i a_i X_i)] \\ &= \Phi_{\underline{X}}(ua_1, \ldots, ua_d) \\ &= \exp(iu(\underline{a}^T \underline{m}) - \frac{1}{2}u^2(\underline{a}^T C \underline{a})) \ . \end{aligned}$$

So $X \sim \mathcal{N}(\underline{a}^T \underline{m}, \underline{a}^T C \underline{a})$, by the characterization of a Gaussian random variable via its characteristic function.

Conversely, if for all $\underline{a} \in \mathbf{R}^d$, $X = \sum_i a_i X_i$ is Gaussian, then in particular each X_i is Gaussian. Hence X_i has a finite mean, say m_i. Also, each X_i has finite variance, and using the Cauchy-Schwarz inequality $E[X_i X_j] \leq (E[X_i^2]E[X_j^2])^{1/2}$, it follows that the covariance matrix of $X_1, X_2, \ldots, X_d$, has finite entries. Call this covariance matrix C. Now, setting with $X = \sum_i u_i X_i$, we see that $E[X] = \underline{u}^T \underline{m}$, and $E[X^2] - E[X]^2 = \underline{u}^T C \underline{u}$. Since X is assumed Gaussian (we assumed that all linear combinations of $X_1, \ldots, X_d$ are Gaussian), we can write

$$\begin{aligned} \Phi_{\underline{X}}(u_1, \ldots, u_d) &= E[\exp(i\sum_i u_i X_i)] \\ &= \Phi_{\sum_i u_i X_i}(1) \\ &= \exp(i\underline{u}^T \underline{m} - \frac{1}{2}\underline{u}^T C \underline{u}) \end{aligned}$$

where in the last step we used the formula for the characteristic function of a Gaussian rv in terms of its mean and variance. But we have now completely determined the joint characteristic function of $X_1, \ldots, X_d$ and, by definition, we see they are jointly Gaussian. □

More generally, any family of random variables arrived at as linear combinations of jointly Gaussian random variables is a jointly Gaussian family of random variables.

Theorem 2 Suppose the real valued random variables $X_1, X_2, \ldots, X_d$ are jointly Gaussian with mean $\underline{m}$ and covariance matrix C. Let $A \in \mathbf{R}^{r \times d}$ and $\underline{b} \in \mathbf{R}^r$. Let $Y_1, \ldots, Y_r$ be defined by $\underline{Y} = A\underline{X} + \underline{b}$. Then $Y_1, \ldots, Y_r$ are jointly Gaussian with mean $A\underline{m} + \underline{b}$ and covariance matrix ACA^T.

Proof

$$\begin{aligned} \Phi_{\underline{Y}}(u_1, \ldots, u_r) &= E[\exp(i\underline{u}^T(A\underline{X} + \underline{b}))] \\ &= \exp(i\underline{u}^T \underline{b})E[i\underline{u}^T A\underline{X}] \\ &= \exp(i\underline{u}^T \underline{b}) \exp(i\underline{u}^T A\underline{m} - \frac{1}{2}\underline{u}^T ACA^T \underline{u}) \\ &= \exp(i\underline{u}^T(\underline{b} + A\underline{m})) \exp(-\frac{1}{2}\underline{u}^T ACA^T \underline{u}) \end{aligned}$$

Conditional expectation for jointly Gaussian random variables

It is very easy to check when a family of jointly Gaussian random variables is mutually independent.

Theorem 3 Let $X_1, X_2, \ldots, X_d$ be real valued random variables that are jointly Gaussian with mean $\underline{m}$ and covariance matrix C. Then $X_1, X_2, \ldots, X_d$ are uncorrelated iff they are independent.

Proof $X_1, X_2, \ldots, X_d$ are uncorrelated iff their covariance matrix C is diagonal. If this is the case, we have

$$\begin{aligned}
\Phi_{\underline{X}}(\underline{u}) &= \exp(i\underline{u}^T \underline{m} - \frac{1}{2}\underline{u}^T C \underline{u}) \\
&= \prod_{k=1}^{d} \exp(iu_k m_k - C_{kk}\frac{u_k^2}{2}) \\
&= \prod_{k=1}^{d} \Phi_{X_k}(u_k) \; .
\end{aligned}$$

But we know that the joint characteristic function of rvs $X_1, X_2, \ldots, X_d$ is separable into their individual characteristic functions iff $X_1, X_2, \ldots, X_d$ are mutually independent.

Conversely, suppose $X_1, X_2, \ldots, X_d$ are mutually independent jointly defined Gaussian rvs. X_i must have mean m_i and variance C_{ii} by assumption. Independence implies the joint characteristic of $X_1, X_2, \ldots, X_d$ is separable into their individual characteristic functions, so we have

$$\Phi_{\underline{X}}(\underline{u}) = \prod_{k=1}^{d} \exp(iu_k m_k - C_{kk}\frac{u_k^2}{2}) \; .$$

But from the form of this joint characteristic function, we see, by definition, that $X_1, X_2, \ldots, X_d$ are jointly Gaussian, and that their covariance matrix is diagonal, i.e. that $X_1, X_2, \ldots, X_d$ are uncorrelated. □

An important consequence of Theorem 1 is the following result :

Theorem 4 Let $X, Y_1, Y_2, \ldots, Y_m$ be jointly Gaussian. Then $E[X \mid \underline{Y}]$ is an affine function of $Y_1, \ldots, Y_d$ (i.e. a constant plus a linear combination of $Y_1, \ldots, Y_d$).

Proof The conditional expectation $E[X \mid \underline{Y}]$ is almost surely uniquely defined as that Borel function of $\underline{Y}$ for which $E[(X - E[X \mid \underline{Y}])g(\underline{Y})] = 0$ for all Borel functions g. In the jointly Gaussian case, it suffices to verify that there is an affine combination $a_0 + \sum_{i=1}^m a_i Y_i$ such that $X - (a_0 + \sum_{i=1}^m a_i Y_i)$ is uncorrelated with the random variables $\underline{Y}$ and has zero mean. This is because, since $(X - (a_0 + \sum_{i=1}^m a_i Y_i), \underline{Y})$ is a linear transformation of $(X, \underline{Y})$, these variables are jointly Gaussian and so this uncorrelatedness would imply that $X - (a_0 + \sum_{i=1}^m a_i Y_i) \amalg \underline{Y}$, which implies that for all Borel functions g

$$E[(X - (a_0 + \sum_{i=1}^{m} a_i Y_i))g(Y_1, \ldots, Y_m)] = E[X - (a_0 + \sum_{i=1}^{m} a_i Y_i)]E[g(Y_1, \ldots, Y_m)] = 0$$

where the second line used $E[X - (a_0 + \sum_{i=1}^m a_i Y_i)] = 0$. This then implies $E[X \mid \underline{Y}] = a_0 + \sum_{i=1}^m a_i Y_i$ by the definition of conditional expectation. Writing down the equations corresponding to this uncorrelatedness and the equation $E[E[X \mid \underline{Y}]] = E[X]$ gives a collection of simultaneous linear equations that can be solved for the coefficients $a_0, a_1, \ldots, a_m$.

You can (and probably should) check that if $(X_1, \ldots, X_n, Y_1, \ldots, Y_m)$ are jointly Gaussian and if Z_i denotes $E[X_i \mid Y_1, \ldots, Y_m]$ for $1 \leq i \leq n$, then $(X_1, \ldots, X_n, Z_1, \ldots, Z_n, Y_1, \ldots, Y_m)$ are jointly Gaussian and the collection of random variables $(X_1 - Z_1, \ldots, X_n - Z_n)$ (which can be thought of as error terms) is independent of $(Y_1, \ldots, Y_m)$.

Example 5 Let X_1, X_2, X_3 be jointly Gaussian with mean $[1, 4, 6]^T$ and covariance matrix $\begin{bmatrix} 3 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix}$. To find $E[X_1 \mid X_2, X_3]$ we write

$$E[X_1 \mid X_2, X_3] = a_0 + a_1(X_2 - 4) + a_2(X_3 - 6) \ .$$

(Note that we already subtracted the means from the conditioning variables to make covariance calculations easier). The equation $E[E[X_1 \mid X_2, X_3]] = E[X_1]$ gives $a_0 = 1$. The requirements that $X_1 - (a_0 + a_1(X_2 - 4) + a_2(X_3 - 6))$ be uncorrelated with X_2 and X_3 respectively give the equations :

$$\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \ .$$

Thus $\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ and

$$E[X_1 \mid X_2, X_3] = 1 + (X_2 - 4) - (X_3 - 6) = X_2 - X_3 + 3 \ .$$

As another example, to find $E[X_2 \mid X_1, X_3]$, we write

$$E[X_2 \mid X_1, X_3] = 4 + b_1(X_1 - 1) + b_2(X_3 - 6) \ .$$

(Note that we have right away observed that the constant term must be the mean of X_2). You can write simultaneous linear equations for b_1 and b_2 based on the requirement that $X_2 - (4 + b_1(X_1 - 1) + b_2(X_3 - 6))$ should be uncorrelated with X_1 and X_3 to conclude that

$$E[X_2 \mid X_1, X_3] = (1/3)X_1 + X_3 - (7/3) \ .$$

Note that $E[X_1 \mid X_3] = E[X_1] = 1$, because X_1 and X_3 are uncorrelated jointly Gaussian rvs, and therefore independent. Using this, we can see from successive conditioning that

$$\begin{aligned} E[X_2 \mid X_3] &= E[E[X_2 \mid X_1, X_3] \mid X_3] \\ &= E[(1/3)X_1 + X_3 - (7/3) \mid X_3] \\ &= 1/3 + X_3 - 7/3 \\ &= X_3 - 2 \ . \end{aligned}$$

This can also be verified directly by solving for c in the equation

$$E[X_2 \mid X_3] = 4 + c(X_3 - 6)$$

by noting that $X_2 - (4 + c(X_3 - 6))$ should be uncorrelated with $(X_3 - 6)$. We get $c = 1$. $\square$