10. Joint Moments and Joint Characteristic Functions

Following section 6, in this section we shall introduce various parameters to compactly represent the information contained in the joint p.d.f of two r.vs. Given two r.vs *X* and *Y* and a function $g(x, y),$ define the r.v

$$Z = g(X, Y) \tag{10-1}$$

Using (6-2), we can define the mean of *Z* to be

$$\mu_Z = E(Z) = \int_{-\infty}^{+\infty} z \, f_Z(z) dz. \tag{10-2}$$

However, the situation here is similar to that in (6-13), and it is possible to express the mean of $Z = g(X,Y)$ in terms of $f_{XY}(x,y)$ *without* computing $f_Z(z)$. To see this, recall from (5-26) and (7-10) that

$$P\left(z < Z \le z + \Delta z\right) = f_Z(z)\Delta z = P\left(z < g(X,Y) \le z + \Delta z\right)$$
$$= \sum \sum_{(x,y)\in D_{\Delta z}} f_{XY}(x,y)\Delta x \Delta y \qquad (10\text{-}3)$$

where $D_{\Delta z}$ is the region in *xy* plane satisfying the above inequality. From (10-3), we get

$$z\, f_Z(z)\Delta z = g(x,y) \sum \sum_{(x,y)\in D_{\Delta z}} f_{XY}(x,y)\Delta x \Delta y. \qquad (10\text{-}4)$$

As Δz covers the entire *z* axis, the corresponding regions $D_{\Delta z}$ are nonoverlapping, and they cover the entire *xy* plane.

By integrating (10-4), we obtain the useful formula

$$E(Z)=\int_{-\infty}^{+\infty} z\, f_Z(z)dz=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)\, f_{XY}(x,y)dxdy. \qquad (10\text{-}5)$$

or

$$E[g(X,Y)]=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)\, f_{XY}(x,y)dxdy. \qquad (10\text{-}6)$$

If X and Y are discrete-type r.vs, then

$$E[g(X,Y)]=\sum_i \sum_j g(x_i,y_j)P(X=x_i, Y=y_j). \qquad (10\text{-}7)$$

Since expectation is a linear operator, we also get

$$E\left(\sum_k a_k g_k(X,Y)\right)=\sum_k a_k E[g_k(X,Y)]. \qquad (10\text{-}8)$$

If X and Y are independent r.vs, it is easy to see that $Z = g(X)$ and $W = h(Y)$ are always independent of each other. In that case using (10-7), we get the interesting result

$$E[g(X)h(Y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x)h(y) f_X(x) f_Y(y)dxdy$$

$$= \int_{-\infty}^{+\infty} g(x) f_X(x)dx \int_{-\infty}^{+\infty} h(y) f_Y(y)dy = E[g(X)]E[h(Y)]. \quad (10\text{-}9)$$

However (10-9) is in general not true (if X and Y are not independent).

In the case of one random variable (see (10- 6)), we defined the parameters mean and variance to represent its average behavior. How does one parametrically represent similar cross-behavior between two random variables? Towards this, we can generalize the variance definition given in (6-16) as shown below:

Covariance: Given any two r.vs *X* and *Y*, define

$$Cov(X,Y) = E\left[(X-\mu_X)(Y-\mu_Y)\right]. \tag{10-10}$$

By expanding and simplifying the right side of (10-10), we also get

$$Cov(X,Y) = E(XY) - \mu_X\mu_Y = E(XY) - E(X)E(Y)$$

$$= \overline{XY} - \overline{X}\,\overline{Y}. \tag{10-11}$$

It is easy to see that

$$\left|Cov(X,Y)\right| \le \sqrt{Var(X)Var(Y)}. \tag{10-12}$$

To see (10-12), let $U = aX + Y,$ so that

$$Var(U) = E\left[\left\{a(X-\mu_X) + (Y-\mu_Y)\right\}^2\right]$$

$$= a^2Var(X) + 2a\,Cov(X,Y) + Var(Y) \ge 0. \tag{10-13}$$

The right side of (10-13) represents a quadratic in the variable *a* that has no distinct real roots (Fig. 10.1). Thus the roots are imaginary (or double) and hence the discriminant

$$\left[Cov(X,Y)\right]^2 - Var(X)\,Var(Y)$$

must be non-positive, and that gives (10-12). Using (10-12), we may define the normalized parameter

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}, \quad -1 \le \rho_{XY} \le 1, \qquad \text{(10-14)}$$

or

$$Cov(X,Y) = \rho_{XY}\,\sigma_X \sigma_Y \qquad \text{(10-15)}$$

and it represents the correlation coefficient between *X* and *Y*.

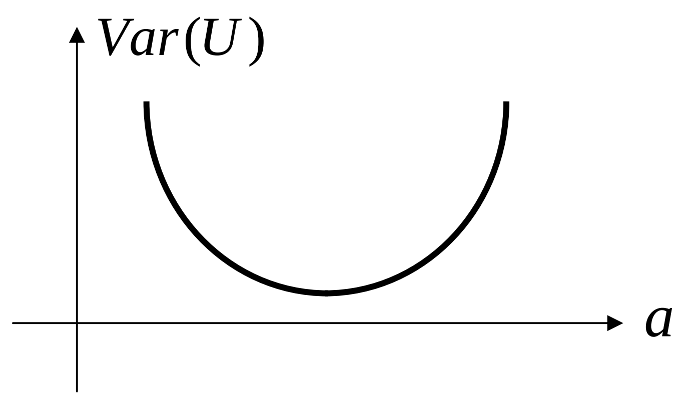

Fig. 10.1

Uncorrelated r.vs: If $\rho_{XY} = 0,$ then X and Y are said to be uncorrelated r.vs. From (11), if X and Y are uncorrelated, then

$$E(XY) = E(X)E(Y). \tag{10-16}$$

Orthogonality: X and Y are said to be orthogonal if

$$E(XY) = 0. \tag{10-17}$$

From (10-16) - (10-17), if either X or Y has zero mean, then orthogonality implies uncorrelatedness also and vice-versa. Suppose X and Y are independent r.vs. Then from (10-9) with $g(X) = X,\ h(Y) = Y,$ we get

$$E(XY) = E(X)E(Y),$$

and together with (10-16), we conclude that the random variables are uncorrelated, thus justifying the original definition in (10-10). Thus independence implies uncorrelatedness.

Naturally, if two random variables are statistically independent, then there cannot be any correlation between them $(\rho_{XY} = 0)$. However, the converse is in general not true. As the next example shows, random variables can be uncorrelated without being independent.

Example 10.1: Let $X \sim U(0,1),\ \ Y \sim U(0,1).$ Suppose *X* and *Y* are independent. Define *Z* = *X* + *Y*, *W* = *X* - *Y* . Show that *Z* and *W* are dependent, but uncorrelated r.vs.

Solution: $z = x + y, \quad w = x - y$ gives the only solution set to be

$$x = \frac{z+w}{2}, \quad y = \frac{z-w}{2}.$$

Moreover $0 < z < 2,\ -1 < w < 1,\ \ z + w \le 2,\ \ z - w \le 2,\ \ z > |w|$
and $|J(z,w)| = 1/2.$

Thus (see the shaded region in Fig. 10.2)

$$f_{ZW}(z,w) = \begin{cases} 1/2, & 0 < z < 2,\ -1 < w < 1,\ z+w \le 2,\ z-w \le 2,\ |w| < z, \\ 0, & \text{otherwise}\ , \end{cases} \tag{10-18}$$

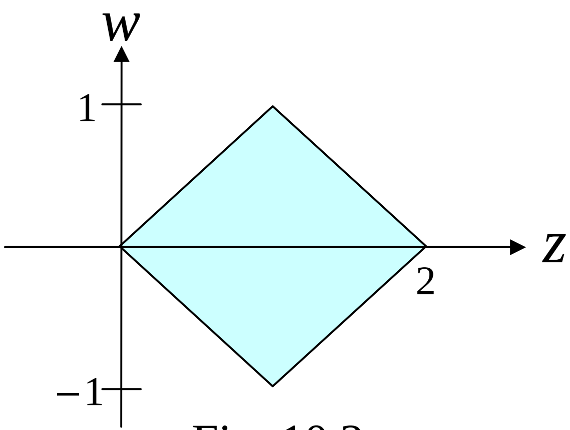

Fig. 10.2

and hence

$$f_Z(z) = \int f_{ZW}(z,w)\,dw = \begin{cases} \int_{-z}^{z} \frac{1}{2}\,dw = z, & 0 < z < 1, \\ \int_{z-2}^{2-z} \frac{1}{2}\,dw = 2 - z, & 1 < z < 2, \end{cases}$$

or by direct computation ($Z = X + Y$)

$$f_Z(z) = f_X(z) \otimes f_Y(z) = \begin{cases} z, & 0 < z < 1, \\ 2 - z, & 1 < z < 2, \\ 0, & \text{otherwise}, \end{cases} \tag{10-19}$$

and

$$f_W(w) = \int f_{ZW}(z, w)dz = \int_{|w|}^{2-|w|} \frac{1}{2}\, dz = \begin{cases} 1 - |w|, & -1 < w < 1, \\ 0, & \text{otherwise}. \end{cases} \tag{10-20}$$

Clearly $f_{ZW}(z, w) \neq f_Z(z) f_W(w)$. Thus Z and W are not independent. However

$$E(ZW) = E\big[(X+Y)(X-Y)\big] = E(X^2) - E(Y^2) = 0, \tag{10-21}$$

and

$$E(W) = E(X - Y) = 0,$$

and hence

$$Cov(Z, W) = E(ZW) - E(Z)E(W) = 0 \tag{10-22}$$

implying that Z and W are uncorrelated random variables.

Example 10.2: Let $Z = aX + bY$. Determine the variance of Z in terms of σ_X, σ_Y and ρ_{XY}.

Solution:

$$\mu_Z = E(Z) = E(aX + bY) = a\mu_X + b\mu_Y$$

and using (10-15)

$$\sigma_Z^2 = Var(Z) = E\left[(Z - \mu_Z)^2\right] = E\left[\left(a(X - \mu_X) + b(Y - \mu_Y)\right)^2\right]$$
$$= a^2 E(X - \mu_X)^2 + 2abE\left((X - \mu_X)(Y - \mu_Y)\right) + b^2 E(Y - \mu_Y)^2$$
$$= a^2\sigma_X^2 + 2ab\rho_{XY}\sigma_X\sigma_Y + b^2\sigma_Y^2. \qquad (10\text{-}23)$$

In particular if X and Y are independent, then $\rho_{XY} = 0,$ and (10-23) reduces to

$$\sigma_Z^2 = a^2\sigma_X^2 + b^2\sigma_Y^2. \qquad (10\text{-}24)$$

Thus the variance of the sum of independent r.vs is the sum of their variances $(a = b = 1)$.

Moments:

$$E[X^k Y^m] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^k y^m f_{XY}(x,y)\,dx\,dy, \qquad (10\text{-}25)$$

represents the joint moment of order (k,m) for X and Y.

Following the one random variable case, we can define the joint characteristic function between two random variables which will turn out to be useful for moment calculations.

Joint characteristic functions:

The joint characteristic function between X and Y is defined as

$$\Phi_{XY}(u,v) = E\left(e^{j(Xu+Yv)}\right) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{j(Xu+Yv)} f_{XY}(x,y)dxdy. \quad (10\text{-}26)$$

Note that $|\Phi_{XY}(u,v)| \le \Phi_{XY}(0,0) = 1.$

It is easy to show that

$$E(XY) = \frac{1}{j^2} \frac{\partial^2 \Phi_{XY}(u,v)}{\partial u \partial v}\bigg|_{u=0,v=0}. \qquad (10\text{-}27)$$

If X and Y are independent r.vs, then from (10-26), we obtain

$$\Phi_{XY}(u,v) = E(e^{juX})E(e^{jvY}) = \Phi_X(u)\Phi_Y(v). \qquad (10\text{-}28)$$

Also

$$\Phi_X(u) = \Phi_{XY}(u,0), \quad \Phi_Y(v) = \Phi_{XY}(0,v). \qquad (10\text{-}29)$$

More on Gaussian r.vs :

From Lecture 7, X and Y are said to be jointly Gaussian as $N(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho),$ if their joint p.d.f has the form in (7-23). In that case, by direct substitution and simplification, we obtain the joint characteristic function of two jointly Gaussian r.vs to be

$$\Phi_{XY}(u,v) = E(e^{j(Xu+Yv)}) = e^{j(\mu_X u+\mu_Y v)-\frac{1}{2}(\sigma_X^2 u^2+2\rho\sigma_X\sigma_Y uv+\sigma_Y^2 v^2)}. \qquad (10\text{-}30)$$

Equation (10-14) can be used to make various conclusions. Letting $v = 0$ in (10-30), we get

$$\Phi_X(u) = \Phi_{XY}(u,0) = e^{j\mu_X u-\frac{1}{2}\sigma_X^2 u^2}, \qquad (10\text{-}31)$$

and it agrees with (6-47).

From (7-23) by direct computation using (10-11), it is easy to show that for two jointly Gaussian random variables

$$Cov(X,Y) = \rho\,\sigma_X\sigma_Y.$$

Hence from (10-14), ρ in $N(\mu_X,\mu_Y,\sigma_X^2,\sigma_Y^2,\rho)$ represents the actual correlation coefficient of the two jointly Gaussian r.vs in (7-23). Notice that $\rho = 0$ implies

$$f_{XY}(X,Y) = f_X(x)\, f_Y(y).$$

Thus if X and Y are jointly Gaussian, uncorrelatedness does imply independence between the two random variables. Gaussian case is the only exception where the two concepts imply each other.

Example 10.3: Let X and Y be jointly Gaussian r.vs with parameters $N(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho).$ Define $Z = aX + bY.$ Determine $f_Z(z).$

Solution: In this case we can make use of characteristic function to solve this problem.

$$\Phi_Z(u) = E(e^{jZu}) = E(e^{j(aX+bY)u}) = E(e^{jauX + jbuY})$$
$$= \Phi_{XY}(au, bu). \qquad (10\text{-}32)$$

From (10-30) with *u* and *v* replaced by *au* and *bu* respectively we get

$$\Phi_Z(u) = e^{j(a\mu_X + b\mu_Y)u - \frac{1}{2}(a^2\sigma_X^2 + 2\rho ab\sigma_X\sigma_Y + b^2\sigma_Y^2)u^2} = e^{j\mu_Z u - \frac{1}{2}\sigma_Z^2 u^2}, \qquad (10\text{-}33)$$

where

$$\mu_Z \triangleq a\mu_X + b\mu_Y, \qquad (10\text{-}34)$$

$$\sigma_Z^2 \triangleq a^2\sigma_X^2 + 2\rho ab\,\sigma_X\sigma_Y + b^2\sigma_Y^2. \qquad (10\text{-}35)$$

Notice that (10-33) has the same form as (10-31), and hence we conclude that $Z = aX + bY$ is also Gaussian with mean and variance as in (10-34) - (10-35), which also agrees with (10-23).

From the previous example, we conclude that any linear combination of jointly Gaussian r.vs generate a Gaussian r.v.

In other words, linearity preserves Gaussianity. We can use the characteristic function relation to conclude an even more general result.

Example 10.4: Suppose X and Y are jointly Gaussian r.vs as in the previous example. Define two linear combinations

$$Z = aX + bY, \quad W = cX + dY. \tag{10-36}$$

what can we say about their joint distribution?
Solution: The characteristic function of Z and W is given by

$$\begin{aligned} \Phi_{ZW}(u,v) &= E(e^{j(Zu+Wv)}) = E(e^{j(aX+bY)u+j(cX+dY)v}) \\ &= E(e^{jX(au+cv)+jY(bu+dv)}) = \Phi_{XY}(au+cv, bu+dv). \end{aligned} \tag{10-37}$$

As before substituting (10-30) into (10-37) with u and v replaced by $au + cv$ and $bu + dv$ respectively, we get

$$\Phi_{ZW}(u,v) = e^{j(\mu_Z u + \mu_W v) - \frac{1}{2}(\sigma_Z^2 u^2 + 2\rho_{ZW}\sigma_X \sigma_Y uv + \sigma_W^2 v^2)}, \tag{10-38}$$

where

$$\mu_Z = a\mu_X + b\mu_Y, \tag{10-39}$$

$$\mu_W = c\mu_X + d\mu_Y, \tag{10-40}$$

$$\sigma_Z^2 = a^2\sigma_X^2 + 2ab\rho\sigma_X\sigma_Y + b^2\sigma_Y^2, \tag{10-41}$$

$$\sigma_W^2 = c^2\sigma_X^2 + 2cd\rho\sigma_X\sigma_Y + d^2\sigma_Y^2, \tag{10-42}$$

and

$$\rho_{ZW} = \frac{ac\sigma_X^2 + (ad+bc)\rho\sigma_X\sigma_Y + bd\sigma_Y^2}{\sigma_Z\sigma_W}. \tag{10-43}$$

From (10-38), we conclude that Z and W are also jointly distributed Gaussian r.vs with means, variances and correlation coefficient as in (10-39) - (10-43).

To summarize, any two linear combinations of jointly Gaussian random variables (independent or dependent) are also jointly Gaussian r.vs.

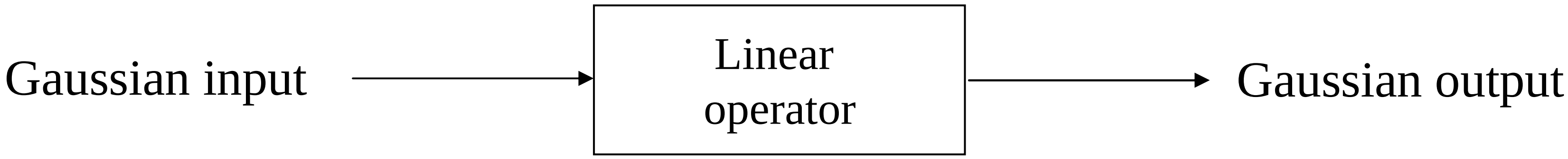

Fig. 10.3

Of course, we could have reached the same conclusion by deriving the joint p.d.f $f_{ZW}(z,w)$ using the technique developed in section 9 (refer (7-29)).

Gaussian random variables are also interesting because of the following result:

Central Limit Theorem: Suppose $X_1, X_2, \cdots, X_n$ are a set of zero mean independent, identically distributed (i.i.d) random variables with some common distribution. Consider their scaled sum

$$Y = \frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}}. \tag{10-44}$$

Then asymptotically (as $n \to \infty$)

$$Y \to N(0, \sigma^2). \tag{10-45}$$

Proof: Although the theorem is true under even more general conditions, we shall prove it here under the independence assumption. Let σ^2 represent their common variance. Since

$$E(X_i) = 0, \tag{10-46}$$

we have

$$Var(X_i) = E(X_i^2) = \sigma^2. \tag{10-47}$$

Consider

$$\Phi_Y(u) = E(e^{jYu}) = E\left(e^{j(X_1+X_2+\cdots+X_n)u/\sqrt{n}}\right) = \prod_{i=1}^{n} E(e^{jX_i u/\sqrt{n}})$$

$$= \prod_{i=1}^{n} \Phi_{X_i}(u/\sqrt{n}) \qquad (10\text{-}48)$$

where we have made use of the independence of the r.vs $X_1, X_2, \cdots, X_n$. But

$$E(e^{jX_i u/\sqrt{n}}) = E\left(1 - \frac{jX_i u}{\sqrt{n}} + \frac{j^2 X_i^2 u^2}{2!\,n} + \frac{j^3 X_i^3 u^3}{3!\,n^{3/2}} + \cdots\right) = 1 - \frac{\sigma^2 u^2}{2n} + o\left(\frac{1}{n^{3/2}}\right), \qquad (10\text{-}49)$$

where we have made use of (10-46) - (10-47). Substituting (10-49) into (10-48), we obtain

$$\Phi_Y(u) = \left[1 - \frac{\sigma^2 u^2}{2n} + o\left(\frac{1}{n^{3/2}}\right)\right]^n, \qquad (10\text{-}50)$$

and as

$$\lim_{n\to\infty} \Phi_Y(u) \to e^{-\sigma^2 u^2/2}, \qquad (10\text{-}51)$$

since

$$\lim_{n\to\infty}\left(1-\frac{x}{n}\right)^n \to e^{-x}. \qquad (10\text{-}52)$$

[Note that $o(1/n^{3/2})$ terms in (10-50) decay faster than $1/n^{3/2}$]. But (10-51) represents the characteristic function of a zero mean normal r.v with variance σ^2 and (10-45) follows.

The central limit theorem states that a large sum of independent random variables each with finite variance tends to behave like a normal random variable. Thus the individual p.d.fs become unimportant to analyze the collective sum behavior. If we model the noise phenomenon as the sum of a large number of independent random variables (eg: electron motion in resistor components), then this theorem allows us to conclude that noise behaves like a Gaussian r.v.

It may be remarked that the finite variance assumption is necessary for the theorem to hold good. To prove its importance, consider the r.vs to be Cauchy distributed, and let

$$Y = \frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}}. \tag{10-53}$$

where each $X_i \sim C(\alpha)$. Then since

$$\Phi_{X_i}(u) = e^{-\alpha|u|}, \tag{10-54}$$

substituting this into (10-48), we get

$$\Phi_Y(u) = \prod_{i=1}^{n} \Phi_X(u/\sqrt{n}) = \left(e^{-\alpha|u|/\sqrt{n}}\right)^n \sim C(\alpha\sqrt{n}), \tag{10-55}$$

which shows that Y is still Cauchy with parameter $\alpha\sqrt{n}$. In other words, central limit theorem doesn't hold good for a set of Cauchy r.vs as their variances are undefined.

Joint characteristic functions are useful in determining the p.d.f of linear combinations of r.vs. For example, with X and Y as independent Poisson r.vs with parameters λ_1 and λ_2 respectively, let

$$Z = X + Y. \tag{10-56}$$

Then

$$\Phi_Z(u) = \Phi_X(u)\Phi_Y(u). \tag{10-57}$$

But from (6-33)

$$\Phi_X(u) = e^{\lambda_1(e^{ju}-1)}, \quad \Phi_Y(u) = e^{\lambda_2(e^{ju}-1)} \tag{10-58}$$

so that

$$\Phi_Z(u) = e^{(\lambda_1+\lambda_2)(e^{ju}-1)} \quad \sim \quad P(\lambda_1 + \lambda_2) \tag{10-59}$$

i.e., sum of independent Poisson r.vs is also a Poisson random variable.