Joint Normal Random Variables

H. Krieger, Mathematics 156, Harvey Mudd College

Fall, 2008

Characteristic Functions: The probabilistic name for the Fourier transform of the distribution of a random variable is the *Characteristic Function.* In other words, if X is a random variable with CDF F_X, then the characteristic function of X, or more properly of F_X, is defined for $t \in \mathbb{R}$ to be

$$\hat{F}_X(t) = E\left(e^{itX}\right) = \int_{-\infty}^{\infty} e^{itx} dF_X(x).$$

Properties: Characteristic functions have similar properties to other transforms, such as generating functions, although some of them are more difficult to prove than for the case of generating functions. For example:

1. The characteristic function of the sum of independent random variables is the product of the characteristic functions of each of the random variables.

2. The characteristic function uniquely determines the distribution.

3. Convergence in distribution is equivalent to pointwise convergence of the corresponding characteristic functions.

4. Change of location and scale: Suppose $\sigma > 0$ and $\mu \in \mathbb{R}$. Then

$$\hat{F}_{\sigma X+\mu}(t) = E\left(e^{it(\sigma X+\mu)}\right) = e^{it\mu}\hat{F}_X(\sigma t).$$

Examples: Here are a few characteristic functions that can be very useful.

1. Suppose X is Bernoulli with probability of success p. Then

$$\hat{F}_X(t) = e^{it}p + (1-p) = 1 - p(1-e^{it}).$$

2. Consequently, if Y is binomial(n, p) then

$$\hat{F}_Y(t) = \left[e^{it}p + (1-p)\right]^n = \left[1 - p(1-e^{it})\right]^n.$$

3. Suppose X is Poisson with parameter $\lambda > 0$. Then

$$\hat{F}_X(t) = e^{-\lambda}e^{\lambda e^{it}} = e^{-\lambda(1-e^{it})}.$$

4. Suppose Z is standard normal, *i.e.* $N(0,1)$, so that Z has density
$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$
for $x \in \mathbb{R}$. Then
$$\hat{F}_Z(t) = e^{-t^2/2} = \sqrt{2\pi}\varphi(t).$$

5. Consequently, if X is normal(μ, σ^2), i.e, $X \sim \sigma Z + \mu$ where $Z \sim N(0,1)$, then
$$\hat{F}_X(t) = e^{it\mu}\hat{F}_Z(\sigma t) = e^{it\mu}e^{-\sigma^2 t^2/2}.$$

Joint Characteristic Functions: If $X = [X_1, X_2, \ldots, X_n]^T$ is a random n-vector, then the characteristic function of X, or of the distribution of X or of the joint distribution of $X_1, X_2, \ldots, X_n$, is defined for $u^T = [u_1, u_2, \ldots, u_n] \in \mathbb{R}^n$ by
$$\hat{F}_X(u) = E\left(e^{iu^T X}\right) = E\left(e^{i(u_1 X_1 + u_2 X_2 + \cdots + u_n X_n)}\right).$$

Properties: Many of the properties of joint characteristic functions are the same. For example:

1. The joint characteristic function of the sum of independent random n-vectors is the product of the joint characteristic functions of each of the random n-vectors.

2. The joint characteristic function uniquely determines the joint distribution.

3. Convergence in distribution is equivalent to pointwise convergence of the corresponding joint characteristic functions.

However, there is one new property, similar to that of joint generating functions, namely: the random variables $X_1, X_2, \ldots, X_n$ are independent *if and only if* the characteristic function of their joint distribution is the product of their individual characteristic functions. In other words, for all $[u_1, u_2, \ldots, u_n] \in \mathbb{R}^n$,
$$\hat{F}_X(u) = E\left(e^{i(u_1 X_1 + u_2 X_2 + \cdots + u_n X_n)}\right) = \prod_{j=1}^{n} E\left(e^{iu_j X_j}\right) = \prod_{j=1}^{n} \hat{F}_{X_j}(u_j).$$

Definition: The random n-vector $X^T = [X_1, X_2, \ldots, X_n]$ is (or the random variables $X_1, X_2, \ldots, X_n$ are) said to have a *joint normal* distribution if and only if the random variable
$$u^T X = u_1 X_1 + u_2 X_2 + \cdots + u_n X_n$$
is normally distributed for every $u^T = [u_1, u_2, \ldots, u_n] \in \mathbb{R}^n$. (Note that we consider constant random variables to be normally distributed with variance 0.)

Theorem: $X_1, X_2, \ldots, X_n$ have a joint normal distribution if and only if the characteristic function of their joint distribution is:

$$\hat{F}_X(u) = e^{iu^T\mu}e^{-u^T\Sigma u/2},$$

where μ is a vector (which is called the mean vector), with components $\mu_j = E(X_j)$, and Σ is a non-negative definite symmetric matrix (which is called the covariance matrix), with entries $\sigma^2_{j,k} = cov(X_j, X_k)$.

Proof: First suppose that $X_1, X_2, \ldots, X_n$ have a joint normal distribution. Then if $u^T = [u_1, u_2, \ldots, u_n] \in \mathbb{R}^n$, we know that the random variable

$$V = u^T X = u_1X_1 + u_2X_2 + \cdots + u_nX_n$$

has a normal distribution. Furthermore,

$$E(u^TX) = u^T\mu \quad \text{and} \quad Var(u^TX) = u^T\Sigma u.$$

Consequently,

$$\hat{F}_X(u) = E\left(e^{i(u^TX)}\right) = \hat{F}_V(1) = e^{iu^T\mu}e^{-u^T\Sigma u/2}.$$

On the other hand, if the characteristic function of their joint distribution is as given and

$$V = u^TX = u_1X_1 + u_2X_2 + \cdots + u_nX_n,$$

then for $t \in \mathbb{R}$ the characteristic function of V must be

$$\hat{F}_V(t) = E(e^{itV}) = E\left(e^{it(u^TX)}\right) = \hat{F}_X(tu) = e^{it(u^T\mu)}e^{-(u^T\Sigma u)t^2/2}.$$

Thus, by the uniqueness of characteristic functions, V must be normally distributed (with mean $u^T\mu$ and variance $u^T\Sigma u$).

Theorem: $X_1, X_2, \ldots, X_n$ have a joint normal distribution with mean vector μ and non-singular (i.e. positive definite) covariance matrix Σ if and only if $X_1, X_2, \ldots, X_n$ have a joint density function given for $x = [x_1, x_2, \ldots, x_n]^T \in \mathbb{R}^n$ by

$$f_X(x) = \frac{1}{(2\pi)^{n/2}(\det\Sigma)^{1/2}}e^{-(x-\mu)^T\Sigma^{-1}(x-\mu)/2}.$$

Proof: Because of the uniqueness property of joint characteristic functions, it suffices to show that for every $u^T = [u_1, u_2, ..., u_n] \in \mathbb{R}^n$, we have

$$\int_{\mathbb{R}^n} e^{iu^Tx}f_X(x)\ dx = e^{iu^T\mu}e^{-u^T\Sigma u/2}.$$

Note that if this is true, then by letting $u = 0 \in \mathbb{R}^n$ we get $\int_{\mathbb{R}^n} f_X(x)\ dx = 1$. Since $f_X(x) \geq 0$ for all $x \in \mathbb{R}^n$, we see that f_X must be a joint density function. In the integral, $\int_{\mathbb{R}^n} e^{iu^Tx}f_X(x)\ dx$, we make a change of variable $y = \Sigma^{-1/2}(x-\mu)$ or, equivalently, $x = \mu + \Sigma^{1/2}y$. This is possible since the positive definite symmetric matrix Σ has a positive definite symmetric square root $\Sigma^{1/2}$ with inverse which we denote by $\Sigma^{-1/2}$. Note that the Jacobian of this transformation

$$\left|\frac{dx}{dy}\right| = \det\left(\Sigma^{1/2}\right) = (\det \Sigma)^{1/2}.$$

Consequently,

$$\begin{aligned}\int_{\mathbb{R}^n} e^{iu^T x} f_X(x)\, dx &= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{iu^T(\mu+\Sigma^{1/2}y)} e^{-y^T y/2}\, dy \\ &= \frac{e^{iu^T\mu}}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{iu^T\Sigma^{1/2}y} e^{-y^T y/2}\, dy.\end{aligned}$$

Letting $v = \Sigma^{1/2}u$, we see that $u^T\Sigma^{1/2}y = v^T y$ so that

$$\begin{aligned}\frac{e^{iu^T\mu}}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{iu^T\Sigma^{1/2}y} e^{-y^T y/2}\, dy &= e^{iu^T\mu} \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{iv_j y_j} e^{-y_j^2/2}\, dy_j \\ &= e^{iu^T\mu} \prod_{j=1}^{n} e^{-v_j^2/2}.\end{aligned}$$

However, $\sum_{j=1}^n v_j^2 = v^T v = u^T \Sigma u$, which gives the result.