EE 464

Spring 2003

Lecture Notes Part 12c

Christopher Wayne Walker, Ph.D.

12.7 Characteristic Functions

Definition: The *characteristic function* of a random vector $\underline{X}$ is

$$\phi(\Omega) = E\left[e^{i\Omega\underline{X}^t}\right] = E\left[e^{i(\omega_1 X_1 + \cdots + \omega_n X_n)}\right]$$

where

$$\underline{X} = [X_1, \ldots, X_n], \quad \Omega = [\omega_1, \ldots, \omega_n]$$

and $\underline{X}^t$ means $\underline{X}$ transpose,

$$\underline{X}^t = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}.$$

Theorem: Suppose X_i are independent random variables with densities $f_i(x_i)$. Let $Z = X_1 + \cdots + X_n$. Then

$$f_Z(z) = f_1(z) * f_2(z) * \cdots * f_n(z).$$

Proof: We will use characteristic functions.

$$E\left[e^{i\omega(X_1+\cdots+X_n)}\right] = E\left[e^{i\omega(X_1}\right] \cdots E\left[e^{i\omega(X_n}\right].$$

This implies

$$\phi_Z(\omega) = \phi_1(\omega)\phi_2(\omega)\cdots\phi_n(\omega)$$

where $\phi_i(\omega)$ is the characteristic function of X_i. Thus,

$$f_Z(z) = f_1(z) * f_2(z) * \cdots * f_n(z).$$

Example: Bernoulli trials and Bernoulli random variables. Say,

$$P(X_i) = p, \; P(X_i = 0) = 1 - p = q.$$

$\{X_i = 1\}$ can correspond to any Bernoulli trial, for example, $X_i = 1$ if the ith toss of a coin is heads.

$$\phi_i(\omega) = E\left[e^{i\omega X_i}\right] = e^{i\omega\cdot 1}p + e^{i\omega\cdot 0}q = pe^{i\omega} + q.$$

Let

$$Z = X_1 + \cdots + X_n.$$

Z can take values $0, 1, 2, \ldots n$.

$$P(Z = k) = P\left(\{X_i = 1\} \text{ occurs exactly } k \text{ times in } n \text{ trials}\right).$$

Now

$$\phi_Z(\omega) = E\left[e^{i\omega Z}\right] = \sum_{k=0}^{n} e^{i\omega k} P(Z = k). \qquad (\star)$$

Also

$$\phi_Z(\omega) = E\left[e^{i\omega(X_1 + \cdots + X_n)}\right] = E\left[e^{i\omega(X_1}\right] \cdots E\left[e^{i\omega(X_n}\right] = \left(pe^{i\omega} + q\right)^n.$$

But the binomial theorem says

$$(a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

Let

$$a = pe^{i\omega}, \quad b = q$$

then

$$\left(pe^{i\omega} + q\right)^n = \sum_{k=0}^{n} \binom{n}{k} p^k e^{i\omega k} q^{n-k} = \sum_{k=0}^{n} e^{i\omega k} P(Z = k)$$

where the last equality follows by $(\star)$. Hence,

$$P(Z = k) = \binom{n}{k} p^k q^{n-k}.$$

So a sum of n independent Bernoulli random variables is binomially distributed.

12.8 Jointly Gaussian Random Variables

We have already given the density function for jointly Gaussian (normal) random variables for two dimensions. Here we just generalize that expression to n dimensions. We get

$$f_{\underline{X}}(\underline{x}) = f(x_1, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}|K|^{1/2}} \exp\left[-\frac{1}{2}(\underline{x} - \underline{\mu})^t K^{-1}(\underline{x} - \underline{\mu})\right]$$

where

$$\underline{x} = (x_1, \ldots x_n)^t, \quad \underline{\mu} = (\mu_1, \ldots \mu_n)^t, \ \mu_i = E[X_i],$$

and K is the $n \times n$ covariance matrix with entries $K_{ij} = Cov(X_i, X_j)$. Note that $Cov(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)]$. We referred to K as C_n earlier in the notes (both conventions are used).

12.9 Central Limit Theorem (CLT)

Theorem(CLT): Let $X_1, \ldots, X_n$ be n independent random variables with finite means μ_i and finite non-zero variances σ_i^2. Let

$$S_n = X_1 + \cdots + X_n,$$

so

$$\mu_S = E(S_n) = \sum_{i=1}^{n} \mu_i, \qquad \sigma_S^2 = Var(S_n) = \sum_{i=1}^{n} \sigma_i^2.$$

Then

$$\lim_{n\to\infty} \frac{S_n - \mu_S}{\sigma_S} \sim N(0,\ 1).$$

We will prove a restricted version of this in the pages to follow.

Let us consider $\overline{X} = S_n/n$. Then

$$\lim_{n\to\infty} \frac{n\overline{X} - \mu_S}{\sigma_S} \sim N(0,\ 1).$$

Now

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

$$E[\overline{X}] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{\mu_S}{n},$$

$$Var(\overline{X}) = \left(\frac{1}{n}\right)^2 \sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}\sigma_S^2.$$

So

$$\mu_S = nE(\overline{X}), \quad \sigma_S^2 = n^2 Var(\overline{X}).$$

Thus,

$$\lim_{n\to\infty} \frac{n\overline{X} - nE(\overline{X})}{n\sqrt{Var(\overline{X})}} \sim N(0,\ 1)$$

or

$$\lim_{n\to\infty} \frac{\overline{X} - E(\overline{X})}{\sqrt{Var(\overline{X})}} \sim N(0,\ 1).$$

Example: Toss a fair coin 100 times. Find the probability you get between 46 and 55 heads.

Solution: We have Bernoulli trials where

$$P(H) = P(T) = 1/2.$$

Let $X_i = 1$ if the ith toss is heads and let $X_i = 0$ if the ith toss is tails. Let

$$S = \sum_{i=1}^{100} X_i.$$

Then, $S \sim B(n,p)$, (Binomial) where $n = 100$, $p = 1/2$. Then $P(46 \text{ to } 55 \text{ H}) = P(45 < S \leq 55)$. Now

$$P(S = k) = \binom{n}{k} p^k q^{n-k} = \binom{100}{k}\left(\frac{1}{2}\right)^{100}.$$

So

$$P(45 < S \leq 55) = \sum_{k=46}^{55} \binom{100}{k}\left(\frac{1}{2}\right)^{100} = 0.680273.$$

The reason for writing $P(45 < S \leq 55)$ instead of $P(46 \leq S \leq 55)$ will become apparent when we use the normal approximation to this problem next.

The CLT allows us to compute the answer approximately using the normal distribution. Even though $n = 100$ is finite let us assume

$$\frac{S - \mu_S}{\sigma_S} \sim N(0,\ 1).$$

Here

$$\mu_S = np = 50, \qquad \sigma_S^2 = npq = 25 \Rightarrow \sigma = 5.$$

Let

$$Z = \frac{S - \mu_S}{\sigma_S}.$$

Then

$$P(45 < S \leq 55) = P\left(\frac{45 - 50}{5} < Z \leq \frac{55 - 50}{5}\right)$$
$$= P(-1 < Z \leq 1) \approx \Phi(1) - \Phi(-1) = 0.682689$$

where,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du.$$

We see that the normal approximation produced a fairly accurate answer in this case.

Restricted Case

Theorem(CLT): Let $f(\cdot)$ be a density with mean μ and variance σ^2. Let $\overline{X}_n$ be the sample mean of n independent random samples of size n from $f(\cdot)$. Let

$$Z_n = \frac{\overline{X}_n - E(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}} = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}.$$

Then,

$$\lim_{n \to \infty} Z_n \sim N(0,\ 1).$$

$\left[\text{Note: } Var(\overline{X}_n) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{\sigma^2}{n}.\right]$

Proof: We will assume the distribution has a moment generating function (mgf). Let

$$M_Z(s) = e^{\frac{1}{2}s^2} \quad \text{(the mgf for } Z \sim N(0,\ 1)\text{)}.$$

Let

$$M_{Z_n}(s) \ \text{ denote the mgf of } Z_n.$$

We will show

$$M_{Z_n}(s) \longrightarrow M_Z(s)$$

which establishes the theorem.

Now

$$\begin{aligned} M_{Z_n}(s) = E\left[e^{sZ_n}\right] &= E\left[\exp\left(s \cdot \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}\right)\right] \\ &= E\left[\exp\left(\frac{s}{n}\sum_{i=1}^{n}\frac{X_i - \mu}{\sigma/\sqrt{n}}\right)\right] \\ &= E\left[\prod_{i=1}^{n}\exp\left(\frac{s}{n}\cdot\frac{X_i - \mu}{\sigma/\sqrt{n}}\right)\right]. \end{aligned}$$

Since the $X_i's$ are independent we get

$$M_{Z_n}(s) = \prod_{i=1}^{n} E\left[\exp\left(\frac{s}{\sqrt{n}}\cdot\frac{X_i - \mu}{\sigma}\right)\right].$$

Let

$$Y_i = \frac{X_i - \mu}{\sigma}.$$

Then

$$M_{Z_n}(s) = \prod_{i=1}^{n} E\left[\exp\left(\frac{s}{\sqrt{n}}\cdot Y_i\right)\right] = \prod_{i=1}^{n} M_{Y_i}\left(s/\sqrt{n}\right).$$

But each X_i has the same density so each Y_i has the same density and therefore the same mgf, call it $M_Y(s)$. Thus

$$M_{Z_n}(s) = \prod_{i=1}^{n} M_Y\left(s/\sqrt{n}\right) = \left[M_Y\left(s/\sqrt{n}\right)\right]^n.$$

The rth derivative of $M_Y\left(s/\sqrt{n}\right)$ evaluated at $s = 0$ gives us the rth moment about the mean of the density $f(\cdot)$ divided by $(\sigma\sqrt{n})^r$. So, with $X = X_i$ we get

$$M_Y\left(s/\sqrt{n}\right) = 1 + \frac{1}{1!}\frac{E(X-\mu)}{\sigma}\frac{s}{\sqrt{n}} + \frac{1}{2!}\frac{E[(X-\mu)^2]}{\sigma^2}\left(\frac{s}{\sqrt{n}}\right)^2$$

$$+\frac{1}{3!}\frac{E[(X-\mu)^3]}{\sigma^3}\left(\frac{s}{\sqrt{n}}\right)^3 + \cdots.$$

But

$$E(X-\mu) = 0, \quad E\left[(X-\mu)^2\right] = \sigma^2.$$

Thus,

$$M_Y\left(s/\sqrt{n}\right) = 1 + \frac{1}{n}\left(\frac{s^2}{2} + \frac{1}{3!}\frac{1}{\sqrt{n}}\frac{E[(X-\mu)^3]}{\sigma^3}s^3 + \frac{1}{4!}\frac{1}{n}\frac{E[(X-\mu)^4]}{\sigma^4}s^4 + \cdots\right).$$

Let

$$u = \frac{s^2}{2} + \frac{1}{3!}\frac{1}{\sqrt{n}}\frac{E[(X-\mu)^3]}{\sigma^3}s^3 + \frac{1}{4!}\frac{1}{n}\frac{E[(X-\mu)^4]}{\sigma^4}s^4 + \cdots.$$

Now

$$\lim_{n\to\infty}\left(1 + \frac{u}{n}\right)^n = e^u \longrightarrow e^{\frac{1}{2}s^2}.$$

Thus

$$\lim_{n\to\infty} M_{Z_n}(s) = M_Z(s) \Rightarrow \lim_{n\to\infty} Z_n \sim N(0,\ 1).$$

12.10 Random Numbers

Often some type of simulation is called for in order to understand how something is going to behave. This can arise for many reasons. For example, a closed form analysis of the problem is not tractable due to the complexity of the mathematics, or perhaps a closed form solution simply does not exist. If the problem at hand has some sort of randomness associated with it, then the analysis becomes even harder. In some cases nice closed form expressions can still be obtained that are either exact or at least a good approximation to the answer. In other cases, however, we will need to resort to a simulation that includes the random effects. To do this we need a random number generator that can produce samples from some desired density function (often a normal density). There is a random number generator called *ran1* in the *Numerical Recipes* book that gives independent *uniformly distributed* random numbers from (0,1). The subroutine *gasdev* in the same book then converts each uniform random number to two independent *normally distributed* random numbers each with mean zero and unit variance (these numbers can then be translated and scaled if other than mean zero unit variance is desired). The routine *gasdev* is a transformation – not a random number generator. The quaility (measure of randomness) of the resulting normal random variables is completely determined by *ran1*. The routine *ran2* is a longer version of *ran1*, i.e., it takes longer for the random number sequence to start repeating itself.

The random number routines and their transformations are available in both FORTRAN and C-code.

12.11 Mean Square Estimation

Definition: The *linear MS estimate* $\hat{S}$ of S in terms of the random variables X_i, is

$$\hat{S} = a_1 X_1 + \cdots a_n X_n,$$

where the a_i are chosen so that

$$e = E\left[(S - \hat{S})^2\right] = E\left[(S - (a_1 X_1 + \cdots a_n X_n))^2\right]$$

is minimized. Now

$$\frac{\partial e}{\partial a_i} = E\left[2(S - (a_1 X_1 + \cdots a_n X_n))X_i\right] = 0$$

for minimal e. Thus,

$$E\left[(S - (a_1 X_1 + \cdots a_n X_n))X_i\right] = 0.$$

This is the orthogonality principle: the error is orthogonal to the data.

We get

$$\begin{array}{ccccccccc} R_{11}a_1 & + & R_{21}a_2 & + & \cdots & + & R_{n1}a_n & = & R_{01} \\ R_{12}a_1 & + & R_{22}a_2 & + & \cdots & + & R_{n2}a_n & = & R_{02} \\ \vdots & & & & & & & & \\ R_{1n}a_1 & + & R_{2n}a_2 & + & \cdots & + & R_{nn}a_n & = & R_{0n} \end{array}$$

where,

$$R_{ij} = E[X_i X_j],\ i,j = 1,2,\ldots,n, \quad R_{0k} = E[SX_k],\ k = 1,2,\ldots,n.$$

Let

$$\underline{X} = [X_1,\ldots,X_n], \quad \underline{A} = [a_1,\ldots,a_n], \quad \underline{R}_0 = [R_{01},\ldots,R_{0n}].$$

Then with

$$\boldsymbol{R} = E[\underline{X}^t \underline{X}]$$

we get (if $\boldsymbol{R}^{-1}$ exists)

$$\underline{A}\boldsymbol{R} = \underline{R}_0 \Rightarrow \underline{A} = \underline{R}_0 \boldsymbol{R}^{-1}.$$

This assumes the data is independent so that $\boldsymbol{R}^{-1}$ exists. If the data is dependent then $\hat{S}$ can be written as a linear combination of a subset of the data with linear independent components.