EE 503

Lecture Notes Part 12

Christopher Wayne Walker, Ph.D.

12.0 Sequences of Random Variables

Here we will deal with n-dimensional versions of some concepts previously studied plus some additional concepts.

12.1 Introduction

Definition: A *random vector* is denoted

$$\underline{X} = [X_1, X_2, \ldots, X_n]$$

where each $X_i,\ i = 1, 2, \ldots, n$ is a random variable.

Given a region D in n-dimensional space, the probability that $\underline{X}$ is in D is given by

$$P(\underline{X} \in D) = \int \cdots \int_D f(\underline{x}) d\underline{x}, \quad \underline{x} = [x_1, x_2, \ldots, x_n].$$

Definition: The *joint distribution* of the random vector $\underline{\mathrm{X}}$ is

$$F(\underline{x}) = F(x_1, x_2, \ldots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n)$$

and the *joint density* is

$$f(\underline{x}) = f(x_1, x_2, \ldots, x_n) = \frac{\partial^n F(x_1, x_2, \ldots, x_n)}{\partial x_1 \partial x_2 \cdots \partial x_n}.$$

Note that integration of $f(x_1, x_2, \ldots, x_n)$ with respect to any of the variables gives the joint density of the remaining variables. For example, if $n = 4$,

$$f(x_1, x_3) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2, x_3, x_4) dx_2 dx_4.$$

12.2 Transformation of a Random Vector

Given $k = n$ functions (if $k < n$ we can use auxillary random variables to make $k = n$ just like we did for $n = 2$ dimensions)

$$g_1(\underline{x}),\ g_2(\underline{x}), \ldots,\ g_n(\underline{x})$$

we form

$$Y_1 = g_1(\underline{X}),\ Y_2 = g_2(\underline{X}), \ldots,\ Y_n = g_n(\underline{X})$$

and solve the system

$$y_1 = g_1(\underline{x}),\ y_2 = g_2(\underline{x}), \ldots,\ y_n = g_n(\underline{x}) \text{ for } \underline{x} = [x_1, x_2, \ldots, x_n].$$

If we have a single solution then

$$f_{\underline{Y}}(y_1, y_2, \ldots, y_n) = \frac{f_{\underline{X}}(x_1, x_2, \ldots, x_n)}{|J(x_1, x_2, \ldots, x_n)|}$$

where

$$J(x_1, x_2, \ldots, x_n) = \begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial y_n}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_n} \end{vmatrix} = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}^{-1}$$

If we have several solutions then we add the corresponding terms as was stated for the 2-dimensional case.

12.3 Independence

Definition: The random variables $X_1, \ldots, X_n$ are called *independent* if each of the events

$$\{\omega : X_1(\omega) \leq x_1\}, \ldots, \{\omega : X_n(\omega) \leq x_n\}$$

are independent.

So $X_1, \ldots, X_n$ independent implies

$$f(x_1, x_2, \ldots, x_n) = f(x_1)f(x_2)\cdots f(x_n),$$

$$f(x_2, x_5) = f(x_2)f(x_5),$$

etc.

Similarly, $X_1, \ldots, X_n$ independent implies

$$F(x_1, x_2, \ldots, x_n) = F(x_1)F(x_2)\cdots F(x_n),$$

$$F(x_2, x_5) = F(x_2)F(x_5),$$

etc.

Furthermore, $X_1, \dots, X_n$ independent implies $g_1(X_1), \dots, g_n(X_n)$ are independent.

In a combined experiment, where we repeat the same underlying experiment over and over independently, then each outcome of the experiment will have the same distribution. In this case we say the random variables formed for each experiment are *independent and identically distributed* (i.i.d).

12.4 Order Statistics

Here we are concerned with an ordered sequence of random variables. Let $X_1, \dots, X_n$ denote n i.i.d. random variables each with cdf $F_X(x)$. Then $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ denotes the ordered version of $X_1, \dots, X_n$, i.e.,

$$X_{(1)} = \min\{X_1, \dots, X_n\}$$

$$X_{(n)} = \max\{X_1, \dots, X_n\}$$

and

$$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}.$$

Notation: In this section, to simplify notation, we will write

$$Y_1 = X_{(1)},\ Y_2 = X_{(2)}, \dots,\ Y_n = X_{(n)}.$$

Theorem: Let $Y_1 \le Y_2 \le \cdots \le Y_n$ represent the order statistics from a cdf $F_X(x)$. Then the marginal cdf of Y_k, $k \in \{1, 2, \dots, n\}$ is

$$F_{Y_k}(y) = \sum_{j=k}^{n} \binom{n}{j} [F_X(y)]^j [1 - F_X(y)]^{n-j}.$$

Proof: Define the indicator function I_L by

$$I_L(x) = \begin{cases} 1, & x \in L, \\ 0, & \text{elsewhere.} \end{cases}$$

Let $y \in \mathbf{R}$ be fixed and let $Z_i = I_{(-\infty,\ y]}(x_i)$. Then

$$\sum_{i=1}^{n} Z_i = \text{number of } X_i \le y.$$

This implies

$$\sum_{i=1}^{n} Z_i \sim B(n, F_X(y)),$$

i.e.,

$$P\left(\sum_{i=1}^{n} Z_i = \beta\right) = \binom{n}{\beta}[F_X(y)]^{\beta}[1 - F_X(y)]^{n-\beta}.$$

Now

$$F_{Y_k}(y) = P(Y_k \leq y)$$

so if $Y_k \leq y$ then the number of $X_i \geq y$ is $\geq k$ and conversely. Thus,

$$F_{Y_k}(y) = P(Y_k \leq y) = P\left(\sum_{i=1}^{n} Z_i \geq k\right)$$

$$= \sum_{j=k}^{n} \binom{n}{j} [F_X(y)]^j [1 - F_X(y)]^{n-j}.$$

Corollary:

$$F_{Y_n}(y) = [F_X(y)]^n$$

and

$$F_{Y_1}(y) = \sum_{j=1}^{n} \binom{n}{j} [F_X(y)]^j [1 - F_X(y)]^{n-j}$$

$$= \sum_{j=0}^{n} \binom{n}{j} [F_X(y)]^j [1 - F_X(y)]^{n-j} - [1 - F_X(y)]^n$$

which implies

$$F_{Y_1}(y) = 1 - [1 - F_X(y)]^n.$$

Theorem: If the X_i are continuous random variables then the density of Y_k (the kth-order statistic) is

$$f_{Y_k}(y) = \frac{n!}{(k-1)!\,(n-k)!} [F_X(y)]^{k-1} [1 - F_X(y)]^{n-k} f_X(y).$$

Proof:

$$f_{Y_k}(y) = \lim_{\Delta y \to 0} \frac{F_{Y_k}(y + \Delta y) - F_{Y_k}(y)}{\Delta y} = \lim_{\Delta y \to 0} \frac{P(y < Y_k \leq y + \Delta y)}{\Delta y}$$

$$= \lim_{\Delta y \to 0} \frac{P\left[(k-1) \text{ of the } X_i \le y, \text{ one } X_i \in (y, y+\Delta y], (n-k) \text{ of the } X_i > y+\Delta y\right]}{\Delta y}$$

$$= \lim_{\Delta y \to 0} \left[\frac{\left[F_X(y)\right]^{k-1} \left[F_X(y+\Delta y) - F_X(y)\right] \left[1 - F_X(y+\Delta y)\right]^{n-k}}{\Delta y} \frac{n!}{(k-1)!\ 1!\ (n-k)!} \right]$$

$$= \frac{n!}{(k-1)!\ (n-k)!} \left[F_X(y)\right]^{k-1} \left[1 - F_X(y)\right]^{n-k} f_X(y).$$

Example: Say $X_i,\ i = 1, 2, \ldots, n$ is exponentially distributed with parameter λ. Then

$$f_{X_i}(x_i) = f_X(x) = \lambda e^{-\lambda x} U(x).$$

Note that

$$E(X_i) = E(X) = \int_0^\infty x \lambda e^{-\lambda x} dx = \frac{1}{\lambda}.$$

Let $Y_1 = \min\{X_1, \ldots, X_n\}$. Then

$$f_{Y_1}(y) = \frac{n!}{(1-1)!\ (n-1)!} \left[F_X(y)\right]^{1-1} \left[1 - F_X(y)\right]^{n-1} f_X(y)$$

$$= n \left[1 - F_X(y)\right]^{n-1} f_X(y).$$

Now

$$F_X(y) = \int_0^y \lambda e^{-\lambda x} dx = \left(1 - e^{-\lambda y}\right) U(y)$$

which implies

$$f_{Y_1}(y) = n \left[e^{-\lambda y}\right]^{n-1} \lambda e^{-\lambda y} U(y)$$

$$= (n\lambda) e^{-(n\lambda) y} U(y).$$

So Y_1 is exponentially distributed with parameter $(n\lambda)$ which implies

$$E(Y_1) = \frac{1}{n\lambda}.$$

Example: Let X be exponentially distributed with parameter λ. Then

$$f_X(x) = \lambda e^{-\lambda x} U(x).$$

Let $Y = cX,\ c > 0$. Then

$$F_Y(y) = P(Y \le y) = P(cX \le y) = P(X \le y/c) = 1 - e^{-\lambda y/c}.$$

Thus

$$f_Y(y) = \frac{\lambda}{c} e^{-\frac{\lambda}{c} y}$$

which implies Y is exponentially distributed with parameter (λ/c). Now let $c = 1/n$. Then Y is exponentially distributed with parameter $(n\lambda)$ which is the same as $Y_1 = \min\{X_1, \ldots, X_n\}$ with each X_i exponentially distributed with parameter λ.

12.5 Mean and Covariance

Theorem: The mean of $g(X_1, X_2, \ldots, X_n)$ is

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \ldots, x_n) f(x_1, x_2, \ldots, x_n) dx_1 \cdots dx_n.$$

Proof: Omitted.

Just like before the covariance of X_i and X_j is

$$C_{ij} = E\left[(X_i - \mu_i)(X_j - \mu_j)\right] = E[X_i X_j] - E[X_i]E[X_j]$$

and

$$\sigma_i^2 = C_{ii}.$$

Definitions: The *sample mean* is given by

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

and the *sample variance* is given by

$$\overline{V} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2.$$

Theorem: If the random variables X_i are uncorrelated with the same mean μ and variance σ^2 then

$$E[\overline{X}] = \mu, \quad E[\overline{V}] = \sigma^2.$$

Proof: Homework problem.

Recall for random variables X_i, X_j,

$$C_{ij} = E\left[(X_i - \mu_i)(X_j - \mu_j)\right] = E[X_i X_j] - E[X_i]E[X_j].$$

This assumes X_i, X_j are real. For complex X_i, X_j,

$$C_{ij} = E\left[(X_i - \mu_i)(X_j^* - \mu_j^*)\right] = E[X_i X_j^*] - E[X_i]E[X_j^*]$$

and

$$\sigma_i^2 = C_{ii} = E[|X_i - \mu_i|^2].$$

We can construct correlation and covariance matrices, respectively, as

$$R_n = \begin{bmatrix} R_{11} & \cdots & R_{1n} \\ \vdots & \vdots & \vdots \\ R_{n1} & \cdots & R_{nn} \end{bmatrix}$$

and

$$C_n = \begin{bmatrix} C_{11} & \cdots & C_{1n} \\ \vdots & \vdots & \vdots \\ C_{n1} & \cdots & C_{nn} \end{bmatrix}$$

where,

$$R_{ij} = E\left[X_i X_j^*\right] = E\left[(X_i^* X_j)^*\right] = \left[E(X_i^* X_j)\right]^* = R_{ji}^*,$$

$$C_{ij} = R_{ij} - \mu_i \mu_j^* = C_{ji}^*.$$

12.6 Conditional Densities

Recall

$$f(y|x) = \frac{f(x,y)}{f(x)}.$$

Similarly, for the n-dimensional case

$$f(x_n, \ldots, x_{k+1} | x_k, \ldots, x_1) = \frac{f(x_1, \ldots, x_n)}{f(x_1, \ldots, x_k)}$$

and

$$F(x_n, \ldots, x_{k+1} | x_k, \ldots, x_1) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_{k+1}} f(\alpha_n, \ldots, \alpha_{k+1} | x_k, \ldots, x_1) d\alpha_{k+1} \cdots d\alpha_n.$$

Now

$$f(x_1, x_2) = f(x_2, x_1) = f(x_2|x_1) f(x_1),$$

$$f(x_1, x_2, x_3) = f(x_3|x_2, x_1) f(x_2, x_1) = f(x_3|x_2, x_1) f(x_2|x_1) f(x_1).$$

Generalizing, we have the chain rule

$$f(x_1, \ldots, x_n) = f(x_n|x_{n-1}, \ldots, x_1) \cdots f(x_2|x_1) f(x_1).$$

Let $n = 3$. Then

$$f(x_1|x_3) = \int_{-\infty}^{\infty} f(x_1, x_2|x_3) dx_2 = \int_{-\infty}^{\infty} f(x_1|x_2, x_3) f(x_2|x_3) dx_2.$$

We also have the concept of conditional expected values.

$$E[X_1|x_2, \ldots, x_n] = \int_{-\infty}^{\infty} x_1 f(x_1|x_2, \ldots, x_n) dx_1.$$

In the above $x_2, \ldots, x_n$ are fixed constants. Thus, $E[X_1|x_2, \ldots, x_n]$ is a constant. Therefore, $E[X_1|X_2, \ldots, X_n]$ is a random variable. We can compute the expected value of this random variable as

$$E\left[E(X_1|X_2, \ldots, X_n)\right] = E(X_1).$$

The proof of this is similar to our proof of

$$E\left[E(X|Y)\right] = E(X).$$

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 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 $P(46 \text{ to } 55 \text{ H}) = P(46 \leq S \leq 55)$. Now

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

So

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

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

$$\begin{aligned} P(46 \leq S \leq 55) &= P\left(\frac{45.5 - 50}{5} < Z \leq \frac{55.5 - 50}{5}\right) \\ &= P(-1 < Z \leq 1) \approx \Phi(1) - \Phi(-1) = 0.680274 \end{aligned}$$

where,

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

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

Note that using 45.5 and 55.5 above in our analysis is called continuity correction. This comes about since $P(X = k)$ for a binomial becomes $P(k - 0.5 \leq X \leq k + 0.5)$ for the normal distribution.

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

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

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

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

12.9 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.

12.10 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,

$$E\left[(S-(a_1X_1+\cdots a_nX_n))X_i\right]=0.$$

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,

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

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.

12.11 Stochastic Convergence

Here we will introduce some convergence concepts.

Consider the infinite sequence of random variables

$$X_1, X_2, X_3, \ldots.$$

Recall X_i is really $X_i(\omega)$, a function of ω.

i. $X_n \to X$ everywhere if $X_n \to X$ as $n \to \infty$.

ii. $X_n \to X$ almost everywhere if $\{\omega \in \Omega : X_n(\omega) \to X(\omega)\}$ exists and has probability 1.

iii. $X_n \to X$ in probability if for any $\epsilon > 0$ $P(|X_n - X| > \epsilon) \to 0$ as $n \to \infty$.

iv. $X_n \to X$ in mean-square sense if $E\left(|X_n - X|^2\right) \to 0$ as $n \to \infty$.

v. $X_n \to X$ in distribution (or Law) if $\lim_{n\to\infty} P\left(X_n \le x\right) = P\left(X \le x\right).$

Convergence in mean-square implies convergence in probability. To see this note by Chebyshev

$$P(|X_n - X| > \epsilon) \le \frac{E\left(|X_n - X|^2\right)}{\epsilon^2} \to 0 \text{ as } n \to \infty.$$

In class we will discuss these concepts, give examples, and show which convergences implies others. We will also study the Borel-Cantelli lemmas.

12.12 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).