CHAPTER 2

PRELIMARIES

2.1 DEFINITION AND WELL-KNOWN RESULTS

The purpose of this chapter is to give necessary background on probability theory needed in this work. By a probability space we mean a triple $(\Omega, \mathfrak{F}, P)$, where

(1) Ω is a set called sample space.

(2) $\mathfrak{F}$ is a Borel field of subsets of Ω,

i.e. $\mathfrak{F}$ is a non-empty family of subsets of Ω such that

(2-1) if $E \in \mathfrak{F}$, then $\bar{E} \in \mathfrak{F}$,

(2-2) if $E_1 \in \mathfrak{F}$ and $E_2 \in \mathfrak{F}$, then $E_1 \cup E_2 \in \mathfrak{F}$,

(2-3) if $E_1, E_2, \ldots$ is a countably infinite sequence of sets belong to $\mathfrak{F}$, then $\bigcup_{n=1}^{\infty} E_n \in \mathfrak{F}$.

(3) P is a probability measure on $\mathfrak{F}$,

i.e. P is a function from $\mathfrak{F}$ into the set R of real number such that

(3-1) For any $E \in \mathfrak{F}$, $P(E) \geq 0$.

(3-2) For any sequence of disjoint sets $E_n \in \mathfrak{F}$, $n = 1, 2, \ldots$, we have

$$P\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} P(E_n) .$$

(3-3) $P(\Omega) = 1.$

Any set in $\mathfrak{F}$ will be referred to as an event. If E is an event, i.e. $E \in \mathfrak{F}$, the value of P at E, i.e. $P(E)$, will be called the probability of the event E.

Now $R^{(k)}$ will be used to denote the set of all k-tuples of real number. If $a = (a_1, \dots, a_k) \in R^{(k)}$, $b = (b_1, \dots, b_k) \in R^{(k)}$ we define $a \leqq b$ to mean $a_i \leqq b_i$ for all i and define $a < b$ to mean $a_i < b_i$ for all i.

The set $\{x \,/\, x \in R^{(k)} \text{ and } a < x \leqq b\}$ will be denoted by $(a,b]$, and the set $\{x \,/\, x \in R^{(k)} \text{ and } x \leqq a\}$ will be denoted by I_a.

By a random vector X we mean a function $X : \Omega \longrightarrow R^{(k)}$ such that for any real number $a = (a_1, \dots, a_k) \in R^{(k)}$, the set $\{w \,/\, w \in \Omega,\ X(w) \in I_a\} \in \mathfrak{F}$. For the case $X : \Omega \longrightarrow R$ we call X a random variable.

For each random vector X we define a function $F_X : R^{(k)} \longrightarrow R$ by $F_X(x) = P(\{w \,/\, X(w) \in I_x\})$. F_X will be called the distribution function of X. Note that if

$$X(w) = (X^{(1)}(w), \dots, X^{(k)}(w)),$$

then each $X^{(i)} : \Omega \longrightarrow R$ is a random variable. Hence $X^{(1)}, \dots, X^{(k)}$ are k random variables. The function F_X is also known as the joint distribution function of the random variables $X^{(1)}, \dots, X^{(k)}$.

If there exists a non-negative Lebesque-measurable function $f : R^{(k)} \longrightarrow R$ such that

$$F(x_1, \ldots, x_k) = \int \underset{I_x}{\cdots} \int f(x_1, \ldots, x_k) dx_1 \ldots dx_k ,$$

we say that f is a density function of F. When this is the case, we have $\frac{\partial^k F(x_1, \ldots, x_k)}{\partial x_k \ldots \partial x_1} = f(x_1, \ldots, x_k)$, at any point where f is continuous.

For any random vector X, the family $\mathfrak{F}'$ of all subsets S' of $R^{(k)}$ such that $X^{-1}(S') \in \mathfrak{F}$ form a Borel field of subsets of $R^{(k)}$. If we define P' by $P'(S') = P(X^{-1}(S'))$ then $(R^{(k)}, \mathfrak{F}', P')$ form a probability space. If $g : R^{(k)} \longrightarrow R$ is any measurable function which is integrable on $S' \in \mathfrak{F}'$ with respect to the measure P' we have

$$\int_S g(X^{(1)}, \ldots, X^{(k)}) d\, P = \int_{S'} g(x_1, \ldots, x_k)\, d\, P' .$$

In case S' is of the form (a,b] and g is Riemann integrable we have

$$\int_{S'} g(x_1, \ldots, x_k) d\, P' = \int_{a_k}^{b_k} \cdots \int_{a_1}^{b_1} g(x_1, \ldots, x_k) dF(x_1, \ldots, x_k).$$

If the integral $\int_\Omega g(X^{(1)}, \ldots, X^{(k)}) dP$ exists, and is finite we define the expectation of g(X), denoted by $E\left[g(X)\right]$, by

$$E\left[g(X)\right] = \int_{\Omega} g(X^{(1)},\ldots, X^{(k)})d\,P\ .$$

Let $g_p(x_1,\ldots, x_k) = x_p$. If $\mu_p = E\left[g_p(X)\right]$ exists for each $p = 1,\ldots, k$, we define $\mu = (\mu_1,\ldots,\mu_p)$ to be the mean vector of X. Furthermore, if $\sigma_{pq} = E\left[(g_p(X)-\mu_p)(g_q(X)-\mu_q)\right]$ exists for each $p, q = 1,\ldots, k$, we define the matrix $\Sigma = (\sigma_{pq})$ to be the covariance matrix of X. In general, if the expectation $E\left[\prod_{j=1}^{k}(g_j(X) - a_j)^{e_j}\right]$ exists, it will be called a product moment of X about $a = (a_1,\ldots, a_k)$ of order $n = e_1+\ldots + e_k$. Note that each σ_{pq} is a product moment about the mean vector of order 2. We shall use the symbol $\mu'_{e_1\ldots e_k}(F)$ to denote the product moment $E(\prod_{j=1}^{k}(g_j(X))^{e_j})$ of any random vector with distribution function F. In the case $k = 1$, $\mu'_e(F)$ denote the $e^{\underline{th}}$ moment about the origin of X.

DEFINITION 2.1.1 A random vector $X = (X_1,\ldots, X_k)$ will be said to have a non-singular normal distribution if it has a density function of the form $f(x_1,\ldots, x_k) = c\ e^{-\frac{1}{2}Q(x_1,\ldots,x_k)}$, where $Q(x_1,\ldots, x_k) = \sum_j \sum_i a_{ij}(x_i - b_i)(x_j - b_j)$ is a definite positive quadratic form.

It can be shown that the mean vector $\mu = (b_1,\ldots, b_k)$, the covariance matrix $\Sigma = (\sigma_{ij})$ is related to $A = (a_{ij})$ by $A = \Sigma^{-1}$

and the constant c is given by $c = \frac{1}{(2\pi)^{\frac{k}{2}}\sqrt{|\Sigma|}}$

Hence the distribution function of non-singular k-variate normal distribution has the density function

$$f(x_1,\ldots,x_k) = \frac{1}{(2\pi)^{\frac{k}{2}}\sqrt{|\Sigma|}} e^{-\frac{1}{2}\sum_j\sum_i \sigma^{ij}(x_i-\mu_i)(x_j-\mu_j)},$$

where σ^{ij} is the ij - entry of Σ^{-1}

Note that for the case of $k = 1$, the density function of a random variable X is $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$, where μ and σ^2 denote the mean and variance of X.

Normal distribution with mean 0 and variance 1 will play an important role in the sequel. We shall refer to it as the <u>standard normal distribution</u>.

<u>DEFINITION 2.1.2</u> The characteristic function of a random vector X is defined as $\varphi(t) = E(e^{it.X}) = \int_{R^{(k)}} e^{it.X}\, dP$, where $i = \sqrt{-1}$, $t \in R^{(k)}$ and $t.X = t_1x_1+\ldots+ t_kx_k$.

It can be shown that the characteristic function of the k-variate normal distribution with the above density function is

$$\varphi(t_1,\ldots,t_k) = \exp(i\sum_{p=1}^{k}\mu_p t_p - \frac{1}{2}\sum_{p,q=1}^{k}\sigma_{pq}\, t_p t_q) .$$

In the case of k = 1, the characteristic function of the standard normal distribution is $\varphi(t) = e^{-\frac{1}{2}t^2}$.
We can obtain the moments of the standard normal distribution by differentiation its characteristic function. We find that its m$^{\underline{th}}$ order moment is given by

$$\mu_m \doteq \begin{cases} 0 & \text{if m is odd integer ,} \\ \dfrac{m!}{2^{\frac{m}{2}}(\frac{m}{2})!} & \text{if m is even integer .} \end{cases}$$

<u>DEFINITION 2.1.3</u> Let $\{X_n\}$ be a sequence of random vectors. If the corresponding sequence $\{F_n\}$ of distribution functions converges to a distribution function F at every continuity point of F, we say that $\{X_n\}$ converges in distribution to F. In this is the case we shall write $X_n \xrightarrow{\mathcal{D}} F$ or $F_n \xrightarrow{\mathcal{D}} F$.

<u>DEFINITION 2.1.4</u> A sequence of distribution function $\{F_n\}$ is said to be convergent, if there is a function F such that

$$\lim_{n\to\infty} F_n(x) = F(x)$$

at every continuity point of F.

We then always find that F is a non-decreasing in each variable x_i and $0 \lneq F(x) \lneq 1$, but it is not necessarily a distribution function. Note that a sequence $\{F_n\}$ may be convergent without converging to a distribution function.

We state without proofs two important theorems on characteristic functions. For their proofs, we refer to [2].

THEOREM 2.1.1 (Uniqueness Theorem)

If X_1, X_2 are random vectors having distribution functions F_1 and F_2 respectively, and characteristic functions φ_1 and φ_2 respectively, a necessary and sufficient condition for $F_1 = F_2$ is that $\varphi_1 = \varphi_2$.

THEOREM 2.1.2 (Continuity Theorem)

Let $\{X_n\}$ be a sequence of random vectors **with corresponding** sequence of characteristic functions $\{\varphi_n\}$. A necessary and sufficient condition for $\{X_n\}$ to converge in distribution to a distribution function F is that the sequence $\{\varphi_n\}$ converges to a limit φ, which is continuous at (0,...,0).

Under these conditions φ is identical with the characteristic function of F.

2.2 RESULTS ON DISTRIBUTIONS FUNCTIONS OF RANDOM VARIABLES

In this section we shall assume that all distribution functions under consideration are distribution functions of random variables.

THEOREM 2.2.1 Any sequence $\{F_n\}$ of distribution functions has a subsequence which converges to a non-decreasing function which is continuous to the right.

Proof Let $\left\{r_n\right\}$ be the enumerable set of all positive and negative rational numbers, including zero, and consider the sequence of real numbers $\left\{F_n(r_1)\right\}$. Since $\left\{F_n(r_1)\right\}$ is bounded, it contains a convergent subsequence. Hence there exists a subsequence $\left\{F_{n'_k}\right\}$ of $\left\{F_n\right\}$ that converges at r_1. By the same argument, we find that $\left\{F_{n'_k}\right\}$ contains a subsequence $\left\{F_{n''_k}\right\}$ which converges at r_1 and r_2. Keep going on, we get $\left\{F_{n'''_k}\right\}$ which is a subsequence of $\left\{F_{n''_k}\right\}$ that converges at r_1, r_2 and r_3. Repeating the same procedure, we obtain successively the subsequence $\left\{F_{n'_k}\right\}$, $\left\{F_{n''_k}\right\}$, $\left\{F_{n'''_k}\right\}$,...., where $\left\{F^{(i)}_{n_k}\right\}$ is a subsequence of $\left\{F^{(i-1)}_{n_k}\right\}$ and $\left\{F^{(i)}_{n_k}\right\}$ converges at $r_1, r_2, \ldots, r_i$.

Define $F_{n_1} = F_{n'_1}$, $F_{n_2} = F_{n''_2}$, $F_{n_3} = F_{n'''_3}$, ..., we see that $\left\{F_{n_k}\right\}$ is a subsequence of $\left\{F_n\right\}$. Put $\lim\limits_{\nu \to \infty} F_{n_\nu}(r_i) = c_i$ for $i = 1, 2, \ldots$, then $\left\{c_i\right\}$ is a bounded sequence, and since every F_{n_ν} is a non-decreasing function, it follows that we have $c_i \leqq c_k$ as soon as $r_i \lneqq r_k$.

Now we define $F(x)$ = greatest lower bound of c_i for all $r_i > x$. It then follows directly from the definition that $F(x)$ is a bounded non-decreasing function. We shall now show that at every continuity point x of F we have

$$\lim_{\nu \to \infty} F_{n_\nu}(x) = F(x) ,$$

so that the subsequence $\left\{F_{n_k}\right\}$ is convergent.

If x is a continuity point of F we can, in fact, choose $h > 0$ such that for any given $\epsilon > 0$, the difference $F(x+h)-F(x-h) < \epsilon$. Let r_i and r_k be rational points such that $r_i \in (x-h,x)$ and $r_k \in (x,x+h)$, so that

$$F(x-h) \leq c_i \leq F(x) \leq c_k \leq F(x+h) .$$

Further, for every ν we have

$$F_{n_\nu}(r_i) \leq F_{n_\nu}(x) \leq F_{n_\nu}(r_k) .$$

As ν tends to infinity, $F_{n_\nu}(r_i)$ and $F_{n_\nu}(r_k)$ tend to the limits c_i and c_k respectively. The difference between these limits is smaller than ϵ, and the quantity $F(x)$ is included between c_i and c_k. Since ϵ is arbitrary, it follows that $\lim_{\nu \to \infty} F_{n_\nu}(x) = F(x)$. Thus the subsequence $\left\{ F_{n_k} \right\}$ is convergent.

<u>THEOREM 2.2.2</u> Let $\left\{ F_n \right\}$ be any sequence of distribution functions. If $\mu_2'(F_n) < k < \infty$ for all n, then any convergence subsequence of $\left\{ F_n \right\}$ converges to a distribution function.

<u>Proof</u> We have, for any $x_o > 0$,

$$K > \mu_2'(F_n) = \int_{-\infty}^{\infty} x^2 dF_n(x) \geq x_o^2 \int_{-\infty}^{-x_o} dF_n(x) + x_o^2 \int_{x_o}^{\infty} dF_n(x).$$

Therefore, we may write

$$\frac{K}{x_o^2} > F_n(-x_o) + 1 - F_n(x_o) , \quad n = 1, 2, \ldots :$$

For a given $\epsilon > 0$, we can therefore choose $x_o > 0$ so that

$$1 - \left[F_n(x) - F_n(-x) \right] < \epsilon \quad \text{for} \quad x > x_o \quad \text{and for all n.}$$

Let $\left\{ F_{n_k} \right\}$ be a convergence subsequence of $\left\{ F_n \right\}$ which converges to a non-decreasing function $G(x)$ at all of its points of continuity. Then clearly for $x > x_o$ we have $1 - \left[G(x) - G(-x) \right] < \epsilon$, that is $\lim_{x \to \infty} \left[G(x) - G(-x) \right] = 1$. Since $G(\infty) - G(-\infty) = 1$ and $G(\infty) \leq 1$, $G(-\infty) \geq 0$. If $G(-\infty) > 0$ or $G(\infty) < 1$ then $G(\infty) - G(-\infty) < 1$, which is a contradiction. Hence $G(-\infty) = 0$ and $G(\infty) = 1$. Therefore $G(x)$ is a distribution function. Hence the convergence subsequence of $\left\{ F_n \right\}$ converges to a distribution function.

The following theorem is an immediate consequence of Theorem 2.2.1 and Theorem 2.2.2.

<u>THEOREM 2.2.3</u> Let $\left\{ F_n \right\}$ be any sequence of distribution functions. If $\mu_2'(F_n) < K < \infty$ for all n, then $\left\{ F_n \right\}$ has a convergence subsequence.

<u>THEOREM 2.2.4</u> Let F, F_n: n = 1, 2,..., be distribution functions such that $\mu_m'(F_n)$, $\mu_m'(F)$ and $\lim_{n \to \infty} \mu_m'(F_n)$ exist. If $\left\{ F_n \right\}$ converges to F, then $\lim_{n \to \infty} \mu_m'(F_n) = \mu_m'(F)$.

<u>Proof</u> Since for any $K > 0$

$$\left| \int_{-\infty}^{\infty} x^m d\, F_n(x) - \int_{-\infty}^{\infty} x^m d\, F(x) \right| \leq A_1 + A_2 + A_3 ,$$

where

$$A_1 = \left| \int_{-K}^{K} x^m d\ F_n(x) - \int_{-K}^{K} x^m d\ F(x) \right| ,$$

$$A_2 = \left| \int_{E_K} x^m d\ F_n(x) \right| ,$$

$$A_3 = \left| \int_{E_K} x^m d\ F(x) \right| ,$$

and E_K is the set of values of x for which $|x| > K$. It follows from Schwarz's inequality that

$$A_2^2 \leqq \int_{E_K} x^{2m} d\ F_n(x). \int_{E_K} d\ F_n(x) ,$$

both integrals being non-negative. Since $\lim_{n \to \infty} \mu'_{2m}(F_n)$ converges, there exists a constant $M_m^2 > 0$ which bounds the first integral for all n and K. Since $F_n \xrightarrow{\mathcal{D}} F$, the second integral on the right, and hence A_2 can be made arbitrary small for all n by choosing K sufficiently large.

Since $\mu'_m(F)$ is finite, A_3 can be made arbitrary small by K sufficiently large.

Since $F_n(x) \xrightarrow{\mathcal{D}} F(x)$ and both $\mu'_m(F_n)$ and $\mu'_m(F)$ are finite, it is evident that for any fixed K, A_1 can be made arbitrary small by choosing n sufficiently large.

$$\text{Hence} \quad \lim_{n\to\infty} \left| \int_{-\infty}^{\infty} x^m d\, F_n(x) - \int_{-\infty}^{\infty} x^m d\, F(x) \right| = 0 ,$$

$$\text{therefore} \quad \lim_{n\to\infty} \mu'_m(F_n) = \mu'_m(F) \quad \text{for each } m.$$

<u>THEOREM 2.2.5</u> Given a sequence of distribution functions $\left\{F_n\right\}$ and a distribution function F. If every convergence subsequence of $\left\{F_n\right\}$ converges to F, then $\left\{F_n\right\}$ converges to F.

<u>Proof</u> Assume the contrary. Hence there exists a continuity point x_o of F such that $\left\{F_n(x_o)\right\}$ does not converge to $F(x_o)$. Therefore there is a subsequence $\left\{F_{n_k}(x_o)\right\}$ of $\left\{F_n(x_o)\right\}$ such that $\lim_{n\to\infty} F_{n_k}(x_o) = 1 \neq F(x_o)$. Since $\left\{F_{n_k}\right\}$ is a subsequence of distribution functions, it has a convergence subsequence. Let $\left\{F_{n_{k_i}}\right\}$ be a convergence subsequence of $\left\{F_{n_k}\right\}$. So $\left\{F_{n_{k_i}}\right\}$ is a convergence sequence but it is also a subsequence of $\left\{F_n\right\}$. Hence $\left\{F_{n_{k_i}}\right\}$ converges to F, therefore $\lim_{n\to\infty} F_{n_{k_i}}(x_o) = F(x_o)$. But $\left\{F_{n_{k_i}}(x_o)\right\}$ is a subsequence of $\left\{F_{n_k}(x_o)\right\}$, which converges to 1. So that $\lim_{n\to\infty} F_{n_{k_i}}(x_o) = 1$. Hence we have $F(x_o) = 1$. This is a contradiction. Hence $\left\{F_n\right\}$ converges to F.

001876

THEOREM 2.2.6 Let F, F_n, n = 1, 2,...., be distribution functions such that $\mu'_m(F)$, $\mu'_m(F_n)$ exist for all n and m. If $\mu'_m(F_n) \xrightarrow{\mathcal{D}} \mu'_m(F)$ for each m then $\{F_n\}$ converges to F.

Proof Since $\{\mu'_2(F_n)\}$ converges, so $\mu'_2(F_n)$ is bounded. By Theorem 2.2.2 we know that every convergence subsequence of $\{F_n\}$ converges to some distribution function.

Let $\{F_{n'_k}\}$, $\{F_{n''_k}\}$ be any two convergent subsequence of $\{F_n\}$. Suppose $\{F_{n'_k}\} \xrightarrow{\mathcal{D}} F'$ and $\{F_{n''_k}\} \xrightarrow{\mathcal{D}} F''$, where F' and F'' are are distribution functions. Hence by Theorem 2.2.4 we obtain

$$\lim_{k\to\infty} \mu'_m(F_{n'_k}) = \mu'_m(F') \text{ and } \lim_{k\to\infty} \mu'_m(F_{n''_k}) = \mu'_m(F'').$$

But from what are given, $\{\mu'_m(F_{n'_k})\}$ and $\{\mu'_m(F_{n''_k})\}$ are subsequences of the same convergence sequence $\{\mu'_m(F_n)\}$, which converges to $\mu'_m(F)$. Hence they converge to the same limit $\mu'_m(F)$, i.e. we have $\mu'_m(F') = \mu'_m(F'') = \mu'_m(F)$, this is true for all m. Therefore F', F'', F have the same characteristic functions, so by Theorem 2.1.1, F' = F'' = F. Hence every convergence subsequence of $\{F_n\}$ converges to the same distribution function F. Hence by Theorem 2.2.5, $\{F_n\}$ converges to F.

COROLLARY 2.1 Let $\left\{ X_n \right\}$ be a sequence of random variables.

If

$$\lim_{n \to \infty} \mu'_m(X_n) = \begin{cases} 0 & \text{when } m \text{ is odd integer,} \\ \dfrac{m}{2^{\frac{m}{2}} \left(\frac{m}{2}\right)!} & \text{when } m \text{ is even integer,} \end{cases}$$

then $\left\{ X_n \right\}$ converges in distribution to the standard normal distribution Φ, where $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} \, dt$.