A PROBABILITY OPERATOR

by

Allan Macdonald Sinclair

Submitted in fulfilment of the requirements for the degree of Master of Science in the Department of Mathematics and Applied Mathematics, Rhodes University, Grahamstown.

Grahamstown.
May, 1964.

INDEX

1 INTRODUCTION

In probability theory it is often convenient to represent laws by characteristic functions, these being particularly suited to classical analysis. Trotter has suggested that probability laws can also be represented by probability operators. These operators are easily handled since they are continuous, and hence bounded, positive linear operators on a normed linear space. This representation arises because distribution functions and their complete convergence correspond to probability operators and their complete convergence. Hence the relations between distribution functions and probability operators will be discussed before the introduction of probability laws.

The first part of section two is devoted to definitions and results of linear spaces. In the second part certain correspondences between distribution functions and probability operators are proved. Rao has proved these for distribution functions of random variables. To extend these results to general distribution functions has involved only minor changes.

In section three necessary and sufficient conditions are given for a linear operator on C (the space of continuous bounded functions on the real line), to be a probability operator. This result is analagous for probability operators to Brochner's Theorem (Loeve, p.207).

The following section depends mainly on theorem 3b. A linear operator A ϵ [C] (the space of all continuous linear operators on C), is shown to be expressible in the form $Ag(y) = \int g(y+x)\, d\mu(x)$, where μ is a function of bounded variation, provided A commutes with every degenerate probability operator.

The last section is independent of the results obtained in sections three and four. It deals pincipally with the relations between probability laws and probability operators. An operation, corresponding to the complex conjugate of a characteristic function, is defined for probability operators. A topology is then introduced into the linear space generated by all probability operators. In this topology, convergence of probability operators corresponds to weak convergence of distribution functions. In this section it is concluded that probability operators are a useful representation of laws, though there are certain disadvantages. An example of their usefulness is Trotter's proof of the classical normal convergence criterion. The main disadvantage is that given a distribution function, the probability operator may not be written down explicitly, as one is often able to do with the characteristic function. Thus characteristic functions have a wider application in probability theory than probability operators have.

The appendix is virtually independent of the other sections. It is shown that a necessary and sufficient condition for a non-degenerate law $\mathcal{L}(X)$ of a random variable X to be $N(o,\sigma^2)$, is that $\mathcal{L}(aX) * \mathcal{L}(bX) = \mathcal{L}(\sqrt{a^2 + b^2}\, X)$ for all a and b≥ o. The method of proof involves characteristic functions.

2 PROBABILITY OPERATORS

Before considering probability operators, some notation of and results of linear spaces are stated. Let R denote the set of real numbers, and C the real linear space of all bounded continuous functions, from R into R, where addition and multiplication by a scalar are defined as usual. A norm is defined for all $g \in C$, by $\|g\| = \sup_{x \in R} |g(x)|$ (Taylor, p. 89). C is then a complete normed linear space, a Banach space or a B-space (Taylor, p.103; D and S, p. 261). The set of all continuous linear operators from C into C is denoted by [C] (Taylor, p. 162). A norm is defined, for all $A \in [C]$, by $\|A\| = \sup \{\|Ag\|: \|g\| \leq 1, g \in C\}$; then [C] is a B-space (Taylor, p.162, D and S, p.61).

Probability operators possess certain ordering properties for which we require some definitions of ordered sets. These are in Taylor on page 391 (or Birkhoff, p.1,16,238). If L is a lattice, we denote sup(a,b) and inf(a,b), where $a \in L$ and $b \in L$, by $a \cup b$ and $a \cap b$. The word positive is used with respect to the zero and partial ordering in the system considered. We now introduce partial orderingsinto C and [C]. Taking "less than or equal to" as the ordering relation in R, we define a partial ordering in C by $f \leqslant g$ iff $f(x) \leqslant g(x)$ for all $x \in R$ and a partial ordering in [C] by $A \leqslant B$

iff $Ag \leqslant Bg$ for all positive g in C. With these definitions C is a vector lattice (the proof is straight forward and is similar to the proof of an example on page 239 in Birkhoff), and [C] is a partially ordered linear space.

Before defining a probability operator one more concept, that of a distribution function, is needed. A distribution function (d.f.), is a non-decreasing function continuous from the left and bounded by 0 and 1 on R. It is not assumed that $F(-\infty) = \lim_{x \to -\infty} F(x) = 0$, nor that $F(+\infty) = \lim_{x \to +\infty} F(x) = 1$. A sequence F_n of d.f.s is said to converge weakly to a d.f. F, written $F_n \xrightarrow{w} F$, iff $F_n(x) \to F(x)$ for all x in the continuity set, C(F), of F. A sequence F_n of d.f.s is said to converge completely, written $F_n \xrightarrow{c} F$, iff $F_n \xrightarrow{w} F$ and $F_n(\pm\infty) \to F(\pm\infty)$ (Loéve, p. 175-178).

We are now in a position to define a probability operator and to investigate some of its properties. If F is a distribution function then the probability operator (pr. op.), T_F is defined for all $y \in R$ and all $g \in C$ by $T_F g(y) = \int g(x+y)dF(x)$. Since g is continuous, the integral may be regarded either as an improper Rieman-Stieltjies integral or as a Lebesgue-Stieltjies integral; the intergral exists since g is bounded. The pr. op. T_F is an operator from C to C:

$$\sup_{y \in R} |T_F g(y)| \leq (F(\infty) - F(-\infty))\|g\| \leq \|g\|,$$ so

$\|T_F g\| \leq \|g\|$; if $z \to y$, then $T_F g(z) \to T_F g(y)$, by the theorem of dominated convergence (Loeve, p.125; Halmos, p.110); thus $T_F g$ is a continuous bounded function on R. Trivially T_F is a linear operator on C. A contraction operator A from C into C is usually defined as an operator for which $\|A\| < 1$. We shall alter this definition slightly to include the case $\|A\| = 1$, so that our definition of a contraction operator A is one for which $\|A\| \leq 1$ (Trotter). With this definition a probability operator is a contraction operator, as is seen from

$$\|T_F\| = \sup_{\|g\| \leqslant 1} \sup_{y \in R} \left| \int g(x+y)\,dF(x) \right|$$

$$\leqslant (F(\infty) - F(-\infty)) \sup_{\|g\| \leqslant 1} \|g\|$$

$$\leqslant 1$$

Since $\|T_F\| < \infty$, T_F is a continuous linear operator on C (Taylor,p.85; D and S.,p.59). Since $T_F g \geqslant 0$, if $g \geqslant 0$, T_F is positive.

a Theorem

There is a one-to-one correspondence between d.f.s (up to additive constants), and pr. op.s.

Proof

If $F_1 = F_2 + C$, where C is a constant, then $T_{F_1} = T_{F_2}$ by definition of T_F.

Let $g_n(x) = 1$ if $x \leqslant -\frac{1}{n}$

$= 0$ if $x \geqslant 0$

$= -nx$ if $-\frac{1}{n} < x < 0$, for $n = 1,2,\ldots,$

then $g_n \in C$. If $T_{F_1} = T_{F_2}$, then

$\lim_n T_{F_1} g_n(-y) = \lim_n T_{F_2} g_n(-y)$, for all $y \in R$.

Since $|g_n| \leqslant 1$, the theorem of dominated convergence applies (Loève,p.125; Halmos,p.110); thus

$\int \lim_n g_n(x-y)dF_1(x) = \int \lim_n g_n(x-y)dF_2(x)$, so that $F_1(y) - F_1(-\infty) = F_2(y) - F_2(-\infty)$ for all $y \in R$.

Otherwise the result follows from the inequality

$$F(y-\tfrac{1}{n}) - F(-\infty)$$
$$\leqslant T_F g_n(-y)$$
$$\leqslant F(y) - F(-\infty)$$ and left continuity of F.

The expression $F[a,b)$, where $a \in R$ and $b \in R$ (or $b = \infty$), will denote $F(b) - F(a)$ (Loève, p.95). This notation arises from the equation

$$F(b) - F(a) = \int_{[a,b)} 1 \; dF(x),$$ where

$[a,b) = \{x : a \leqslant x < b\}$. $F(-\infty,b)$ will denote $F(b) - F(-\infty)$.

b   <u>Theorem</u>

A sequence F_n of d.f.s converges completely (up to additive constants), to a d.f. F iff $T_{F_n} g(y) \to T_F g(y)$ for all $y \in R$ and all $g \in C$.

<u>Proof</u>

If $F_n \xrightarrow{c} F$ (up to additive constants), then, by the Helly-Bray Theorem (Loeve, p.182), since

$g(x+y) \in C$, it follows that $T_{F_n} g(y) \to T_F g(y)$ for all $y \in R$.

By the definition of the g_n,

$$F(y-\tfrac{1}{n}) - F(-\infty) = \int_{-\infty}^{y-\frac{1}{n}} dF(x)$$

$$\leq T_F g_n(-y)$$

$$\leq \int_{-\infty}^{y} dF(x) = F(y) - F(-\infty), \text{ for } y \in R \text{ and } n = 1,2\ldots$$

Similarly $F_k(y-\tfrac{1}{n}) - F_k(-\infty)$

$$\leq T_{F_k} g_n(-y)$$

$$\leq F_k(y) - F_k(-\infty).$$

Therefore

$$F_k(-\infty,\ y-\tfrac{1}{n}) - F(-\infty,\ y)$$

$$\leq T_{F_k} g_n(-y) - T_F g_n(-y)$$

$$\leq F_k(-\infty,\ y) - F(-\infty,\ y-\tfrac{1}{n}).$$

Now if $y \in C(F)$ then, given $\epsilon>0$, there is $\delta = \delta(\epsilon, y)$ such that

$F(-\infty,\ y+\tfrac{1}{n}) - F(-\infty,\ y-\tfrac{1}{n}) < \epsilon$ for $\tfrac{1}{n} < \delta$.

Fix $y \in C(F)$ and fix $n > \dfrac{1}{\delta(\epsilon, y)}$.

Then for $k \geq K = K(\epsilon, y, n)$, $|T_{F_k} g_n(-y) - T_F g_n(-y)| < \epsilon$.

From this inequality and a previous pair of inequalities it follows that

$F_k(-\infty, y-\frac{1}{n}) - F(-\infty, y) < \epsilon$ and

$F_k(-\infty, y) - F(-\infty, y-\frac{1}{n}) > -\epsilon$, for $k \geqslant K$.

In the first of these inequalities replace y by $y + \frac{1}{n}$; then, for $k \geqslant K_1 = K(\epsilon, y+\frac{1}{n}, n)$

$F_k(-\infty, y) - F(-\infty, y+\frac{1}{n}) < \epsilon$.

Now for $k \geqslant \max(K, K_1)$,

$$F_k(-\infty, y) - F(-\infty, y)$$

$$> F(-\infty, y-\frac{1}{n}) - \epsilon - F(-\infty, y)$$

$$> -2\epsilon \text{ and}$$

$$F_k(-\infty, y) - F(-\infty, y)$$

$$< \epsilon + F(-\infty, y+\frac{1}{n}) - F(-\infty, y) < 2\epsilon.$$

Therefore, for $k \geqslant \max(K, K_1)$,

$|F_k(-\infty, y) - F(-\infty, y)| < 2\epsilon$.

Thus $F_k \overset{H}{\to} F$ (up to additive constants).

$(T_{F_k} - T_F)\ 1(y)$

$= \int 1\ dF_k(x) - \int 1\ dF(x)$

$= F_k(-\infty, +\infty) - F(-\infty, +\infty)$

$\to 0$ as $k \to \infty$, so, by a theorem in Loève (p.178), $F_k \overset{c}{\to} F$ (up to additive constants).

The above result is, as noted by Rao, a special case of Alexandroff's theorem (D. and S.,p.316). Trotter proved the second part of this for a continuous limit d.f, and Rao proved it when the d.f.s are d.f.s of r.v.s . Let M denote the set of all non-increasing continuous functions bounded by o and 1 on R. The above shows that $F_k \overset{c}{\to} F$ (up to additive constants)

$\Rightarrow T_{F_k} g(y) \to T_F g(y),\ g \in C.$

$\Rightarrow T_{F_k} g(y) \to T_F g(y),\ g \in M,$

$\Rightarrow F_k \overset{c}{\to} F$ (up to additive constants).

A sequence of pr. op.s T_{F_k} is said to converge completely iff

$T_{F_k} g(y) - T_F g(y) \to 0$ for all $y \in R$ and all $g \in C$.

Theorem

If F_1 and F_2 are d.f.s, then $T_{F_1} T_{F_2} = T_{F_2} T_{F_1} = T_{F_1 * F_2}$

$(F_1(-\infty) = F_2(-\infty) = 0)$

Proof

$(T_{F_1} T_{F_2} g)(y)$

$= T_{F_1} \int g(y+z)\, dF_2(z)$

$= \int \left\{ \int g(y+x+z)\, dF_2(z) \right\} dF_1(x)$

$= \int \left\{ \int g(y+x+z)\, dF_1(x) \right\} dF_2(z)$, by Fubini's Theorem,

$= T_{F_2} \int g(y+x)\, dF_1(x)$

$= (T_{F_2} T_{F_1} g)(y)$, for all $y \in R$ and for all $g \in C$.

The use of Fubini's Theorem is justified:

$g(y+x+z)$ is a continuous function of x and z, since g is continuous;

$\iint |g(y+x+z)|\, dF_1(x)\, dF_2(z) \leq \|g\|$. (Loeve, p.136; Halmos, p. 147).

A function F on R is said to be composed of d.f.s F_1 and F_2 (or is the convolution of F_1 and F_2), written $F = F_1 * F_2$, iff
$F(y) = \int F_1(y-x)\, dF_2(x)$, $y \in R$, where it is assumed that $F_1(-\infty) = F_2(-\infty) = 0$; then $F = F_1 * F_2 = F_2 * F_1$ (Loeve, p.193).

Let $F(y) = F_1 * F_2(y) = \int F_1(y-x)\, dF_2(x)$,

let $a < b$ and let $a = x_{n1} < \ldots\ldots < x_{n,k_n+1} = b$

with $\sup_k (x_{n,k+1} - x_{nk}) \to 0$ as $n \to \infty$.

Then for every $g \in C$ and $y \in R$,

$$\int_a^b g(y+x)\, dF(x)$$

$$= \lim_n \sum_k g(y+x_{nk})\, F[x_{nk}, x_{n,k+1})$$

$$= \lim_n \sum_k g(y+x_{nk}) \int F_1[x_{nk} - u, x_{n,k+1} - u)\, dF_2(u)$$

$$= \iint_a^b g(y+x)\, dF_1(x-u)\, dF_2(u)$$

$$= \iint_{a-u}^{b-u} g(y+x+u)\, dF_1(x)\, dF_2(u)$$

Therefore, letting $a \to -\infty$ and $b \to +\infty$,

$T_F g(y) = T_{F_1} T_{F_2} g(y)$, by Fubini's Theorem.

The second part of the proof is similar to the proof of the composition theorem (Loève, p.193).

3 A CONDITION FOR A LINEAR OPERATOR TO BE A PROBABILITY OPERATOR

The following lemma is a convenient tool in the theorem of this section, since we may restrict our attention from C to the class M of all non-increasing continuous functions bounded by 0 and 1.

a <u>Lemma</u>

If two continuous linear operators, A and B, on C are equal on the class of all non-increasing continuous functions bounded by 0 and 1, then, A = B on C.

<u>Proof</u>

By hypothesis, A = B on M.

Since every function in the set, M_1 of all non-decreasing continuous functions bounded by 0 and 1 is a difference of two members of M, A = B on M_1.

Let $\epsilon > 0$ be arbitrary but fixed, such that $\epsilon \leqslant 4$; let g be any fixed member of C. Then we shall construct a set of functions $h_k \epsilon M_1 \cup M$ and a set of constants a_k ($|a_k| \leqslant 1$), $k = \pm 1, \pm 2, \ldots\ldots\ldots$, such

that $\|g - g(0) - \sum_{k=-\infty}^{\infty} a_k h_k\| < \epsilon$.

Since g is bounded and continuous on R, it is uniformly continuous on every finite interval [a, b]. Thus there exists $\delta_1 = \delta_1(\epsilon) = \frac{1}{n_1}$ such that, on

$[0, 1]$, $|g(x) - g(y)| < \frac{\epsilon}{4}$ for $|x - y| < 2\delta_1$.

Then if $0 \leqslant x \leqslant \delta_1$

$|g(x) - \{g(0) + h_1(x)[g(\delta_1) - g(0)]\}| < \frac{\epsilon}{2}$,

where $h_1(x) \epsilon M_1$ such that

$h_1(x) = 0$ for $x \leqslant 0$

$\quad = 1$ for $x \geqslant \delta_1$; let $a_1 = g(\delta_1) - g(0)$, then

$|a_1| < \frac{\epsilon}{4} \leqslant 1$. Let $h_k(x) \epsilon M_1$ be chosen so that

$h_k(x) = 0$ for $x \leqslant (k-1)\delta_1$

$\quad = 1$ for $x \geqslant k\delta_1$; let

$a_k = g(k\delta_1) - g(\overline{k-1}\delta_1)$, for $k = 1, \ldots\ldots\ldots n_1$.

Then if $0 \leqslant x \leqslant n_1\delta_1 = 1$,

$$|g(x) - g(0) - \sum_{k=1}^{n_1} a_k h_k(x)| < \frac{\epsilon}{2}.$$

This method may be applied to the intervals [1, 2], [2, 3],......... in succession. Since [0, ∞) is the union of an enumerable number of these intervals each of which adds on a finite number of h_k's, it follows that there is a sequence $\{h_k\}$ of elements of M_1 and a sequence $\{a_k\}$, $|a_k| \leqslant 1$, of constants such that

$$|g(x) - g(0) - \sum_1^{\infty} a_k h_k(x)| < \frac{\epsilon}{2} \text{ for } x \geqslant 0.$$

The series terminates for fixed finite x and is thus convergent for each $x \in R$. The same type of process may now be applied to $(-\infty, 0]$, using M in place of M_1. Thus

$$|g(x) - g(0) - \sum_{k=-\infty}^{-1} a_k h_k(x)| < \frac{\epsilon}{2} \text{ for } x \leqslant 0,$$

where $h_k \in M$ and $|a_k| \leqslant 1$.

The series $\sum_{k=-\infty}^{-1} a_k h_k(x) = 0$ if $x \geqslant 0$

and $\sum_{k=1}^{\infty} a_k h_k(x) = 0$ if $x \leqslant 0$.

Thus $\|g - g(0) - \sum_{k=-\infty}^{\infty} a_k h_k(x)\| \leqslant \frac{\epsilon}{2}$, where a_k are constants ($|a_k| \leqslant 1$) and the h_k are suitably chosen members of $M_1 \cup M$. $\sum_{k=-\infty}^{\infty} a_k h_k(x)$ is a continuous function of x and is bounded by $\frac{\epsilon}{2} + 2\|g\|$; $\sum_k a_k h_k$ is a member of C.

Therefore

$$\|Ag - Bg\|$$

$$\leqslant (\|A\| + \|B\|)\epsilon + \|A(g(0)) - B(g(0))\|$$

$$+ \sum_k \|Ah_k - Bh_k\|$$

$\leqslant (\|A\| + \|B\|)\epsilon$, since A = B on $M \cup M_1$.

Since ϵ is arbitrary and $\|A\| + \|B\|$ is finite, $Ag = Bg$ for all $g \in C$. Thus A = B on C; the proof is complete.

If $F_a(x) = 1$ for $x > a$

$= 0$ for $x \leqslant a$, then F_a is called a degenerate d.f. (Loève, p.202). The pr. op. corresponding to F_a is

$$T_{F_a}\, g(y) = \int g(y + x)\, dF_a(x)$$
$$= g(y + a).$$

The linear operator T_a, $a \in R$, defined, from C onto C, by $T_a g(x) = g(a + x)$ is called a degenerate pr. op..

a <u>Theorem</u>

A necessary and sufficient condition for a linear operator T, from C into C, to be a pr. op. is that T be a positive contraction operator that commutes with every degenerate pr. op.

<u>Proof</u>

If F is a d.f. then $\|T_F\| \leqslant 1$, T_F is positive and $T_F T_a = T_a T_F$; the necessity is proved.

To prove the sufficiency, a d.f. F corresponding to T is defined, then T_F is proved to be equal to T on M; the result then follows by lemma 3a.

Let $g_{ny} = 1$ if $z \leqslant y - \frac{1}{n}$

$= 0$ if $z \geqslant y$

$= -n(z - y)$ if $y - \frac{1}{n} < z < y$, for $n = 1, 2,$ ...... and all $y \in R$. Then $0 \leqslant Tg_{ny} \leqslant 1$ and,

if $n_1 > n_2$ then, $g_{n_1 y} \geqslant g_{n_2 y}$, so

$Tg_{n_1 y}(z) \geqslant Tg_{n_2 y}(z)$, for all y and $z \in R$.

Thus $\lim_n Tg_{ny}(z)$ exists for all y and $z \in R$.

Let $F_0(y) = \lim_n Tg_{ny}(0)$. If $y_1 \geqslant y_2$ then

$g_{ny_1} \geqslant g_{ny_2}$, so

$$F_0(y_1) = \lim_n Tg_{ny_1}(0)$$

$$\geqslant \lim_n Tg_{ny_2}(0) = F_0(y_2).$$

Obviously $0 \leqslant F_0(y) \leqslant 1$.

Let $F(y) = F_0(y)$ if y is a point of left continuity of F_0,

$= F_0(y - 0)$ if y is a point of left discontinuity of F_0.

Then F is a d.f. (Loève, p.175), and, if $C(F_o)$ is the continuity set of F_o in R then, $F = F_o$ on $C(F_o)$. $C(F_o)$ is dense in R: $\overline{C(F_o)} = R$.

Now $(Tg_{ny})(x)$

$= (T_xT)\{g_{ny}(z)\}(0)$, by definition of T_x,

$= (TT_x)\{g_{ny}(z)\}(0)$, by hypothesis,

$= T\{g_{ny}(z + x)\}(0)$

$= T\{g_{n,y-x}(z)\}(0)$, since $z + x \leqslant y \equiv z \leqslant y - x$

$= Tg_{n,y-x}(0)$; a similar result holds with T_F in place of T.

By definition of F,

$\lim_n T\,g_{ny}(0) = F(y)$, if $y \in C(F_o)$; also

$\lim_n T_F g_{ny}(0) = F(y)$, for all $y \in R$ (l. c).

Thus $\lim_n (T - T_F)\, g_{ny}(0) = 0$, if $y \in C(F_o)$.

Therefore $\lim_n (T - T_F)\, g_{ny}(x) = 0$, if $y \in x + C(F_o)$, where $x + C(F_o)$ denotes the set $\{x + z : z \in C(F_o)\}$.

Let I be any set dense in R. Let g be a fixed member in M, then, given $\epsilon > 0$, there exist $y_k \in I$, $k = 1, 2, \ldots\ldots K = K(\epsilon, g)$, and $N = N(\epsilon, g)$ such that

$$\|g - g(\infty) - \sum_{k=1}^{K} \epsilon g_{ny_k}\| \leqslant \epsilon, \text{ for all } n \geqslant N.$$

Let $x_1, \ldots\ldots x_K$ be such that

$g(x_1) - g(\infty) = \epsilon$, $g(x_2) - g(x_1) = \epsilon, \ldots\ldots$

$g(x_K) - g(x_{K-1}) = \epsilon$; all the x_r are assumed to be finite, thus $g(-\infty) \leqslant K\epsilon + g(\infty)$.

$x_1 > x_2 > \ldots\ldots > x_K$.

Let $\delta = \delta(\epsilon, g) = \frac{1}{2} \min_{2 \leqslant r \leqslant K} (x_r - x_{r-1})$ and let $N = N(\epsilon, g) = [\frac{1}{\delta}] + 1$. Let $y_k \in I$, $k = 1, \ldots K$, be such that $x_k + \delta < y_k < x_k + 2\delta$.

Let $G(x) = g(x) - g(\infty) - \sum_{k=1}^{K} \epsilon g_{ny_k}(x)$, where

$n \geq N$. Then $|G(x)| \leq \epsilon$ for $x > x_1$,
$G(x_1) = 0$, $|G(x)| \leq \epsilon$ for $x_1 > x > x_2$, $G(x_2) = 0$,
and so on. Thus $|G(x)| \leq \epsilon$ for all $x \in R$.

Therefore $\|g(x) - g(\infty) - \sum_{k=1}^{K} \epsilon g_{ny_k}(x)\| \leq \epsilon$ for

$n \geq N$.

There is an $x \in R$ such that
$\|Tg - T_F g\| < |Tg(x) - T_F g(x)| + \epsilon$, by definition
of the norm.

By the approximation to g,

$$\|Tg - T_F g\|$$

$$< |T\{g - g(\infty) - \sum_{k=1}^{K} \epsilon g_{ny_k}\}(x)|$$

$$+ |T_F\{g - g(\infty) - \sum_{k=1}^{K} \epsilon g_{ny_k}\}(x)|$$

$$+ |(T - T_F)1| \cdot g(\infty)$$

$$+ \epsilon \sum_{k=1}^{K} |(T - T_F)\, g_{ny_k}(x)| + \epsilon$$

CORRECTIONS ON PAGE 24

Insert the following between " ... $T = T_F$ on M_o . " (L.7) and "The proof is complete ."(L.8).

If we now prove that $T1 = T_F 1$, then it will follow that $T = T_F$ on M , from the above inequalities, since

$$(T - T_F)1(x) \, . \, g(\infty) = 0 \quad .$$

Let

$$\begin{aligned} g_k(x) &= 1 && k + 1/4 \leqslant x \leqslant k + 3/4 \\ &= 0 && x \leqslant k - 1/4 \text{ or } x \geqslant k + 5/4 \\ & \text{linear } k - 1/4 \text{ to } k + 1/4 \text{ from } 0 \text{ to } 1 \\ & \text{linear } k + 3/4 \text{ to } k + 5/4 \text{ from } 1 \text{ to } 0 . \end{aligned}$$

Then $1(x) = \sum_{k=-\infty}^{\infty} g_k(x)$ and g_k is the difference of two members of M_o , so that $(T - T_F)\, g_k = 0$

Therefore

$$\begin{aligned} & \|(T - T_F)1\| \\ \leqslant & \sum_k \|(T - T_F)\, g_k\| \\ = & \ 0 \quad . \end{aligned}$$

$$\leqslant 3\epsilon + \epsilon \sum_{k=1}^{K} |(T - T_F)\, g_{ny_k}(x)|, \text{ for } n \geqslant N, \text{ if } g \in M_0 = \{g : g \in M \text{ and } g(\infty) = 0\}.$$

Therefore

$\|Tg - T_F g\|$

$$\leqslant 3\epsilon + \epsilon \sum_{k=1}^{K} |\lim_{n} (T - T_F)\, g_{ny_k}(x)|.$$

If the dense set I is chosen to be $x + C(F_0)$, then,

$\lim_{n} |(T - T_F)\, g_{ny_k}(x)| = 0.$

Therefore $\|Tg - T_F g\| \leqslant 3\epsilon$; hence $T = T_F$ on M.

The proof is complete.

The above theorem corresponds, for pr. op.s to Bochner's Theorem (Loève, p.207), for ch. f.s. The following lemma permits us to reduce the conditions in theorem 3b.

c <u>Lemma</u>

If $T \in [C]$ commutes with T_x for every $x \in H$, which is a set of real numbers that generates an additive group dense in R then T commutes with every degenerate probability operator.

Proof

If x and $y \epsilon H$, then

$$T_{-x}T = T_{-x}TT_xT_{-x}$$

$$= T_{-x}T_xTT_{-x}$$

$$= TT_{-x} \quad \text{and}$$

$$T_{x+y}T = T_xT_yT$$

$$= TT_xT_y$$

$$= TT_{x+y}$$

Thus T commutes with T_x, where x is in the additive group generated by H.

If $TT_xg = T_xTg$ for every $g \epsilon M$, then, by lemma 3a, $TT_x = T_xT$. Fix $x \epsilon R$ and $g \epsilon M$, then g is uniformly continuous on R. Let x_n be a sequence of real numbers, in the additive group generated by H, tending to x.

$$|T_x Tg(y) - TT_x(y)|$$

$$\leqslant |T_x Tg(y) - T_{x_n} Tg(y)| + |TT_{x_n} g(y) - TT_x g(y)|$$

$$= |Tg(y + x) - Tg(y + x_n)| + \left|T\Big(g(x_n + z) - g(x + z)\Big)(y)\right|$$

$\to 0$ as $x_n \to x$, since Tg is continuous and

$$\sup_{z \in R} |g(x_n + z) - g(x + z)| \to 0 \text{ as } x_n \to x.$$

The proof is complete.

Using this lemma the statement of theorem 3b. may be reduced, since it is only required that $TT_x = T_x T$ for all $x \in H$.

4 REPRESENTATION OF A SPECIAL TYPE OF LINEAR OPERATOR

We have defined a partial ordering in [C] by $A \leqslant B$ iff $Af \leqslant Bf$ for all $f \in C$ and $0 \leqslant f$, and have noted that C is a partially ordered vector space. It can be shown that [C] is not a vector lattice (see appendix B). This difficulty is not serious, as the linear operators that are considered form a lattice. Before this is proved some preliminary definitions and results will be given.

Let BD denote the B-space of all bounded functions on R with norm $\|g\| = \sup_{x \in R} |g(x)|$.

[C] is a closed linear subspace of BD. For $A \in [C]$ define A^+ as follows:

if $0 \leqslant f \in C$,

$A^+f(x) = \sup\{Ag(x) : 0 \leqslant g \leqslant f;\ g \in C\}$, and for general $f \in C$, $A^+f = A^+(f \cup 0) - A^+(-(f \cap 0))$.

A^+ is a linear operator from C into BD. If $0 \leqslant f$, $|A^+f(x)| \leqslant \|A\|\,\|f\|$ and so $\|A^+f\| \leqslant \|A\|\,\|f\|$. In general $\|A^+f\| \leqslant \|A^+(f \cup 0)\| + \|A^+(-(f \cap 0))\|$

$$\leqslant \|A\|\,\|f\| + \|A\|\,\|f\|$$

$$\leqslant 2\|A\|\,\|f\|.$$

From this and by definition of A^+f it follows that A^+ is an operator from C into BD. The result that $A^+cf = cA^+f$, where $c \in R$ and $f \in C$, will follow in general if it holds for $c \geq 0$ and $0 \leq f$; the latter follows from

$$A^+(cf)(x) = \sup\{Acg(x) : 0 \leq cg \leq cf;\ cg \in C\}$$
$$= c \sup\{Ag(x) : 0 \leq g \leq f;\ g \in C\}$$
$$= cA^+f(x).$$

It only remains to prove that $A^+(f + h) = A^+f + A^+h$, for all f and $h \in C$. This will follow in general if it holds for $0 \leq f$ and $0 \leq h$. The following inequalities prove the particular case:

$$A^+(f + h)(x) \geq \sup\{A(g + k)(x) : 0 \leq g \leq f,\ 0 \leq k \leq h\}$$
$$= A^+f(x) + A^+h(x) \quad \text{and}$$
$$A^+(f + h)(x) = \sup\{Ag(x) : 0 \leq g \leq f + h\}$$
$$= \sup\{A(g \cap f)(x) + A(g \cap h)(x) : 0 \leq g \leq f+h\}$$
$$\leq A^+f(x) + A^+h(x)$$

If $0 \leq f$, then $A^+f(x) \geq 0$.

Now by the above A^+ will be a member of [C] iff $A^+(C) \subset C$. This will follow iff A^+f is continuous for every f such that $0 \leq f \in C$. We prove that this holds iff A^+1 is continuous.

Let $\epsilon > 0$ be fixed but arbitrary. Let $f \geq 0$ be a fixed member of C and x a fixed member of R. If $\delta > 0$, denote $\{z : |z - x| < \delta\}$ by $K(x, \delta)$. There is a $g \in C$ such that $0 \leq g \leq f$ and

$A^+f(x) < Ag(x) + \frac{\epsilon}{2}$ and a $\delta_1 = \delta_1(x, \epsilon, g)$ such that

$|Ag(x) - Ag(y)| < \frac{\epsilon}{2}$ for $y \in K(x, \delta_1)$.

Then $A^+f(x) - A^+f(y)$

$< Ag(x) - Ag(y) + \frac{\epsilon}{2}$

$< \epsilon$, for $y \in K(x, \delta_1)$.

Similarly there exists a $\delta_2 > 0$ such that

$A^+h(x) - A^+h(y) < \frac{\epsilon}{2}$, for $y \in K(x, \delta_2)$, where

$h = \|f\| - f$.

Since $f = \|f\| - h$,

$A^+f(x) - A^+f(y)$

$= \|f\|\{A^+1(x) - A^+1(y)\} - \{A^+h(x) - A^+h(y)\}$

$> \|f\|\{A^+1(x) - A^+1(y)\} - \frac{\epsilon}{2}$ for $y \in N(x, \delta_2)$. Thus

A^+f will be continuous at x if there is a $\delta_3 > 0$

such that $A^+1(x) - A^+1(y) > -\frac{\epsilon}{2\|f\|}$ for $y \in K(x, \delta_2)$.
The latter will hold by the above working iff A^+1 is continuous at x.

Theorem

If $A \in [C]$ commutes with every degenerate pr. op. there is a function μ, of bounded variation on R, such that

$Ag(y) = \int g(x + y)\, d\mu(x)$, for all $g \in C$ and all $y \in R$.

Proof

We prove, firstly, that if A commutes with every degenerate pr. op. then $A^+ \in [C]$. By a previous remark, to prove this we have only to show that A^+1 is continuous.

If $0 \leqslant f \in C$, then

$A^+T_yf(x)$

$= \sup\{AT_yg(x) : 0 \leqslant T_yg \leqslant T_yf,\ T_yg \in C\}$

$= \sup\{T_yAg(x) : 0 \leqslant g \leqslant f,\ g \in C\}$,

since $T_yC = C$ and $T_yg \leqslant T_yf \Leftrightarrow g \leqslant f$,

$= \sup \{Ag(x+y) : 0 \leq g \leq f,\ g \in C\}$

$= A^+f(x+y)$

$= T_y A^+ f(x)$. Applying this to 1 we obtain

$A^+1(y) = A^+T_y1(0)$

$\quad = A^+(T_y1)(0)$

$\quad = A^+1(0)$ for all $y \in R$. This implies that A^+1 is a constant and is thus continuous.

Let $A^- = A^+ - A$, then $A^- \in C$.

If $0 \leq f$, then $Af \leq A^+f$ so that $0 \leq A^-f$. Since $A^+T_y = T_yA^+$, for all $y \in R$ on the set of positive functions in C, it follows that A^+, and hence A^-, commutes with every degenerate pr. op. Thus $\frac{A^+}{\|A^+\|}$ and $\frac{A^-}{\|A^-\|}$ are positive contraction operators that commute with every degenerate pr. op. They satisfy the conditions of the theorem in section 3, so there exist d.f.s F_+ and F_- such that

$\frac{A^+}{\|A^+\|} g(y) = \int g(x+y)\, dF_+(x)$ and

$\frac{A^-}{\|A^-\|} g(y) = \int g(x+y)\, dF_-(x)$.

Therefore $Ag(y) = \int g(x + y)\, d\mu(x)$,

where $\mu(x) = \|A^+\| F_+(x) - \|A^-\| F_-(x)$.

Since μ is the difference of two monotonic bounded functions it is a function of bounded variation (D. and S., p.350; Rudin, p.102).

The comment at the end of lemma 3c is applicable to the above theorem.

5 PROBABILITY CONCEPTS

A probability space $(\Omega, \mathcal{Q}, P)$, is a normed measure space; Ω is the space, $\mathcal{Q}$ is a σ-algebra over Ω and P is a measure defined on $\mathcal{Q}$ such that $P(\Omega) = 1$ (Loeve, p. 150; Halmos, p.191). Throughout the following discussion it will be assumed that the pr. space is fixed. A random variable (r. v.) is a finite measurable function from Ω to R, where the σ-field in R is the Borel σ-field B. Corresponding to the r.v. X, on the pr. space $(\Omega, \mathcal{Q}, P)$, there is a non-negative set function P_X, called the distribution of X, defined by $P_X S = P\{x : x \in \Omega, X(x) \in S\}$ for $S \in B$. P_X is a pr. on the σ-algebra B; hence X generates, on its range space, a pr. space (R, B, P_X) (Loève, p.166). To the pr. P_X on B there corresponds a unique d.f. F_X defined on R by $F_X(x) = P_X(-\infty, x) = P\{y : y \in \Omega, X(y) < x\}$. A sequence X_n of r.v.s is said to converge in pr. to a r.v. X, written $X_n \xrightarrow{P} X$, iff $P\{x : x \in \Omega, |X_n(x) - X(x)| \geq \epsilon\} \to 0$, for every $\epsilon \to 0$. The characteristic function (ch. f.) of a d.f.F is defined by $f(u) = \int e^{iux}\, dF(x)$, $u \in R$ (Loeve, p.185).

There is a one-one correspondence between distributions, d.f.s (up to additive constants), ch.f.s and pr. op.s . The one-one correspondence between distributions and d.f.s (up to additive constants), is established in Loeve on page 96; between d.f.s (up to additive constants) and ch.f.s follows from the inversion formula for ch.f.s (Loeve, p.186), and between d.f.s (up to additive constants) and pr.op.s is proved in 1.a. Thus distributions, d.f.s (up to additive constants), ch.f.s and pr.op.s represent the same mathematical concept, called a pr. law (or law). A law is denoted by $\mathcal{L}$ and if the d.f., corresponding to the law, is the d.f. of a r.v.X then the law is written $\mathcal{L}(X)$. This view is strengthened by the following correspondences:

$F_1 * F_2 = F \leftrightarrow f_1 f_2 = f \leftrightarrow T_{F_1} T_{F_2} = T_F$ (Loeve, p.193),

where corresponding subscripts denote corresponding ch.f.s and d.f.s. The law $\mathcal{L}$, corresponding to the ch.f.f, is said to be composed of the laws $\mathcal{L}_1$ and $\mathcal{L}_2$, corresponding to the ch.f.s f_1 and f_2, denoted by $\mathcal{L} = \mathcal{L}_1 * \mathcal{L}_2$, iff $f = f_1 f_2$. Complete convergence of d.f.s up to additive constants is equivalent to pointwise convergence of the corresponding ch.f.s and to complete convergence of the corresponding pr. op.s. A sequence $\mathcal{L}_n$ of laws, corresponding to d.f.s F_n, is said to converge to a law $\mathcal{L}$, corresponding to a d.f. F, iff $F_n \xrightarrow{c} F$ up to additive constants.

The above paragraph may be summarised as follows: if corresponding laws, d.f.s, ch.f.s and pr.op.s have the same subscript, then

$$\begin{aligned} &\mathcal{L}_1 = \mathcal{L}_2 \\ \Leftrightarrow\ & F_1 = F_2 \text{ (up to additive constants)} \\ \Leftrightarrow\ & f_1 = f_2 \\ \Leftrightarrow\ & T_1 = T_2\,, \end{aligned}$$

$$\begin{aligned} &\mathcal{L} = \mathcal{L}_1 * \mathcal{L}_2 \\ \Leftrightarrow\ & F_1 = F_1 * F_2 \text{ (where } F_1(-\infty) = F_2(-\infty) = 0) \\ \Leftrightarrow\ & f = f_1 f_2 \\ \Leftrightarrow\ & T = T_1 T_2 \text{ and} \end{aligned}$$

$\mathcal{L}_n \to \mathcal{L}$

$\Leftrightarrow F_n \overset{c}{\to} F$ (up to additive constants),

$\Leftrightarrow f_n \to f$

$\Leftrightarrow T_n g(y) \to Tg(y)$, for all $y \in R$ and all $g \in C$.

The usefulness of ch.f.s in the representation of pr. laws is that ch.f.s are ideally suited to classical analysis; ch. f. s are continuous bounded complex functions of a real variable u; the complex conjugate of a ch.f. of a r.v.X corresponds to the r.v.-X; ~~a ch.f. of~~ a r.v. is symmetric iff the ch.f is real (a r.v. is called symmetric iff $P\{y : y \in \Omega, X(y) < x\} = P\{y : y \in \Omega, X(y) > -x\}$); pointwise convergence of ch.f.s corresponds to convergence of their laws; pointwise convergence of the indefinite integrals of ch.f.s corresponds to weak convergence of the d.f.s (Loeve, p.190). Thus the usual operations, on the complex functions of a real variable, applied to a ch.f. have corresponding operations of interest in pr. theory. Another advantage of ch.f.s is given a particular d.f. (e.g. Normal), the corresponding ch.f. may often be written down as an explicit function in terms of elementary functions; this is important in applications of pr. theory to statistics.

Just as ch.f.s are suited to classical analysis so pr.op.s are suited to the theory of linear operators. Pr.op.s are bounded (and hence continuous) linear operators defined on a B-space; the real B-algebra generated by the set of all pr.op.s is commutative; complete convergence of pr.op.s is equivalent to convergence of their laws. An operation on pr.op.s corresponding to the complex conjugate of a ch.f. has not been given. The linear operator $\bar{T}_F$ is defined for all $g \in C$, by $\bar{T}_F g(y) = \int g(y - x) dF(x)$. $\bar{T}_F$ is a pr. op. and $\bar{T}_X = T_{-X}$. The d.f. of $-X$ is $1 - F_X(-x + 0)$, thus

$$\begin{aligned} \bar{T}_X g(y) &= \int g(y - x)\, dF_X(x) \\ &= -\int g(y + x)\, dF_X(-x) \\ &= \int g(y + x)\, dF_{-X}(x), \text{ since} \end{aligned}$$

$F(-x + 0) = F(-x)$ on $C(F)$, so that $\bar{T}_X g = T_{-X} g$;

alternatively a proof may be given using expectations (Loeve, p.149).

$\bar{T}_X g(y) = Eg(y - X) = T_{-X} g(y)$.

X is a symmetric r.v. iff $T_X = \bar{T}_X$;

$\mathcal{L}(X) = \mathcal{L}(-X) \Leftrightarrow T_X = T_{-X} \Leftrightarrow T_X = \bar{T}_X$.

The operation T_F to $\bar{T}_F$ on pr.op.s is continuous.

Given $\epsilon > 0$, let $\|T_{F_1} - T_{F_2}\| < \frac{\epsilon}{2}$, then

$\|\bar{T}_{F_1} - \bar{T}_{F_2}\|$

$\leqslant \|\bar{T}_{F_1} g - \bar{T}_{F_2} g\| + \frac{\epsilon}{4}$, for some $g :: \|g\| \leqslant 1$,

$\leqslant |\bar{T}_{F_1} g(y) - \bar{T}_{F_2} g(y)| + \frac{\epsilon}{2}$, for some $y \in R$,

$= |T_1 f(0) - T_2 f(0)| + \frac{\epsilon}{2}$, where $f(x) = g(y - x)$.

$\leqslant \|T_1 - T_2\| + \frac{\epsilon}{2} < \epsilon$.

The use of probability operators in pr. theory is best illustrated by Trotter's proof (Trotter) of the classical normal convergence criterion (or classical central limit theorem). Rao uses pr.op.s - if $E \in B$ then $T_X^E g(y) = \int_E g(x + y)\, dF(x)$, where F is the d.f. corresponding to the r.v.X - to prove a theorem that is a generalisation of the classical normal convergence criterion (Rao). Pr.op.s provide another tool in probability theory that can be used to complement ch.f.s; each used for a purpose suited to its self. There is one big drawback to pr.op.s that does not occur with ch.f.s: a pr.op. of a given d.f. may not be written down explicitly in the same way that the ch.f. may be (e.g. the ch.f. of the law N(0, 1) is $e^{\frac{-u^2}{2}}$).

Weak convergence of d.f.s may be related to a form of convergence of pr.op.s. The topology in which convergence of pr.op.s corresponds to weak convergence of d.f.s must be weaker than the topology in which the d.f.s converge completely, since complete convergence of d.f.s implies weak convergence of d.f.s

The first criterion for weak convergence is suggested by a remark made by Loève (Loève, p.372), Let C_0 be the subspace of C, whose members vanish at infinity - thus if $g \in C_0$ $\lim_{x \to \pm\infty} g(x) = 0$. The following result holds:

$F_n \overset{W}{\to} F$(up to additive constants) if, and only if, $T_{F_n} g(y) \to T_F g(y)$ for all $y \in R$ and all $g \in C_0$.

The proof is similar to that of result 2b. Instead of using the Helly-Bray Theorem (Loeve p.182), one uses the extended Helly-Bray lemma (Loeve p. 181). In place of the g_n's defined in 2b one uses a function with two steps so that it is zero outside some interval.

An alternative approach to weak convergence may be obtained by using linear functionals. If V is a linear space then V^f denotes the space of all linear functionals on V. A subset Γ of V^f is called total if, and only if, $f(x) = 0$ for all $f \in \Gamma$, implies $x = 0$ (Taylor, p.153). Let Φ denote the linear space generated by all pr.op.s— $\Phi \subset [C]$. If $T \in \Phi$ then there is a unique function μ continuous on the left and of bounded variation on R with $\mu(-\infty) = 0$ such that
$Tg(y) = \int g(x + y)\, d\mu(x)$ for all $y \in R$ and $g \in C$. Φ is a B-space.

Define L_y by $L_y T = \mu(y)$ where μ is the function of bounded variation that corresponds to $T = T_\mu \in \Phi$. Then $L_y \in \Phi^f$, since

$L_y(aT_\mu + bT_\nu)$
$= L_y(T_{a\mu + b\nu})$
$= a\mu(y) + b\nu(y)$
$= aL_yT_\mu + bL_yT_\nu$

Let $C(T)$ denote the continuity set of the function of bounded variation μ corresponding to $T \in \Phi$; then $C(T)$ is dense in R. Let $\Gamma(T)$ denote $\{L_y : y \in C(T)\}$. If $L_yT_\nu = 0$ for all $y \in C(T)$, then $\nu(y) = 0$ for all $y \in C(T)$ and since ν is continuous on the left $\nu = 0$. Therefore $T_\nu = 0$ and so $\Gamma(T)$ is a total subset of Φ^f. The set $\Gamma(T)$ may be used to define a topology in Φ, called the $\Gamma(T)$ topology of Φ (Taylor, p. 151 - 153; Dunford and Schwartz, p.419). A sequence T_n of elements in Φ converges in the $\Gamma(T)$ topology to T_o iff $L_yT_n \to L_yT_o$ for all $L_y \in \Gamma(T)$.

Weak convergence of d.f.s (up to additive constants) to a d.f. F is equivalent to convergence of the corresponding pr.op.s to a pr.op. T_F, in the $\Gamma(T_F)$ topology. There is no loss in generality in assuming $F(-\infty) = 0$: a change of the additive constants will ensure this.

$F_n \overset{W}{\to} F$ (up to additive constants) $\Leftrightarrow F_n + c_n \overset{W}{\to} F$, where c_n are constants. By a result in Loeve on page 178, $F_n + c_n \overset{W}{\to} F$ and $F(-\infty) = 0$ imply $\lim_n F_n(-\infty) = 0$. Thus the c_n are assumed to be chosen so that $F_n(-\infty) + c_n = 0$.

Therefore $F_n \overset{W}{\to} F$ (up to additive constants)

$\Leftrightarrow L_y T_{F_n} + c_n \to L_y T_F$, if $y \in C(F)$

$\Leftrightarrow T_{F_n} + c_n \to T_F$ in the $\Gamma(T_F)$ topology of Φ

$\Leftrightarrow T_{F_n} \to T_F$ in the $\Gamma(T_F)$ topology of Φ.

Appendix A

A CONDITION THAT A LAW BE NORMAL

A law is called normal, denoted by $N(b, \alpha^2)$, if the ch.f. is $f(u) = \exp[iub - \frac{\alpha^2 u^2}{2}]$(Loeve, p.215).

Trotter notes in his paper on the classical normal convergence criterion that the only properties of the normal law used in the proof are $aN(0, \alpha^2) = N(0, a^2\alpha^2)$ and $N(0, \alpha^2) * N(0, \beta^2) = N(0, \alpha^2 + \beta^2)$. Thus these conditions must be sufficient to characterise the normal law; in this section this is proved by ch.f.s .

A necessary and sufficient condition that $\mathcal{L}(X)$ not degenerate at zero be $N(0, \alpha^2)$, for some real $\alpha > 0$, is that $\mathcal{L}(aX) * \mathcal{L}(bX) = \mathcal{L}(\sqrt{a^2 + b^2}\, X)$, for all real a and $b \geqslant 0$.

Proof

If $\mathcal{L}(X)$ is $N(0, \alpha^2)$ then $f(u) = e^{\frac{-u^2\alpha^2}{2}}$

CORRECTIONS ON PAGES 44 AND 45

Delete " $H(s) = e^{\alpha s}$ for some constant α ....

.... for all $x \geqslant 0$."

(From L.15, p.44 to L.7, p.45)

Insert " $H(s) = e^{cs}$ for some complex constant c (p.96, Aczél, J. , Vorlesungen über Funktionalgleichungen und ihre Anwendungen ,1961) . "

NOTE ON APPENDIX A

After the thesis had been typed the author's attention was drawn to Aczél page 96 where the same result is proved .

and the ch.f. corresponding to

$\mathcal{L}(aX) * \mathcal{L}(bX)$ is

$$f(au).f(bu) = e^{-u^2(a^2 + b^2)\frac{\alpha^2}{2}}$$

$$= f(\sqrt{a^2 + b^2}\ u)$$

Now assume that

$\mathcal{L}(aX) * \mathcal{L}(bX) = \mathcal{L}(\sqrt{a^2 + b^2}\ X)$ for all $a \geq 0$, $b \geq 0$. Let $f(u)$ be the ch.f. of $\mathcal{L}(X)$. Then

$f(au).f(bu) = f(\sqrt{a^2 + b^2}\ u)$ (Loeve, p.194).

Let $H(x^2) = f(x)$; hence

$H(a^2u^2).H(b^2u^2) = H(a^2u^2 + b^2u^2)$, and let

$a^2u^2 = s$, $b^2u^2 = t$.

Therefore $H(s)H(t) = H(s + t)$ for all s and $t \geq 0$,

$H(0) = f(0) = 1$, since $F(0) = F(\infty) - F(-\infty)$, and H is continuous since f is.

The solution of this equation is well known:

~~$H(s) = e^{\alpha s}$ for some constant α (Rudin, Ex.3, p.162).~~

~~Let $H(1) = \beta$, then by induction~~

~~$H(\frac{1}{n}) = \beta^{\frac{1}{n}}$, $n = 1, 2, \ldots\ldots$~~

~~and $H(\frac{m}{n}) = \beta^{\frac{m}{n}}$, $m = 1, 2, \ldots$~~

~~$n = 1, 2, \ldots$~~

~~Thus $\beta \neq 0$, since if this were the case,~~

~~$H(0) = 0$, by continuity; contradiction. In the~~

~~same way β is finite (or otherwise use $\|H\| \leqslant 1$).~~

~~Thus let c be such that $e^{c} = \beta$, then c is finite.~~

~~Then $H(\frac{m}{n}) = e^{\frac{m}{n}c}$; let $x \geqslant 0$ then, given $\epsilon > 0$, there is~~

~~a $\delta = \delta(\epsilon, x)$ such that $|e^{xc} - e^{yc}| < \frac{\epsilon}{2}$ and~~

~~$|H(x) - H(y)| < \frac{\epsilon}{2}$ for $|y - x| < \delta$.~~

~~Then $|H(x) - e^{cx}| < \epsilon$, for $|\frac{m}{n} - x| < \delta$~~

~~so that $H(x) = e^{cx}$ for all $x \geqslant 0$.~~

Therefore $f(u) = H(u^2) = e^{cu^2}$ for all real u.

Since f is a ch.f. $f(-u) = \overline{f(u)}$, so

$e^{cu^2} = e^{\bar{c}u^2}$ for all real u.

Therefore $c = \bar{c}$ and so c is real.

$|f(u)| \leqslant 1$, thus $c \leqslant 0$, for if $c > 0$ then $e^{c} > 1$.

If $c = 0$ then $f(u) = 1$ for all $u \in R$.

This implies that X degenerates at zero (Loeve, p.202).

Thus $c < 0$, say $c = -\frac{\alpha^2}{2}$, where $\alpha > 0$; then

$f(u) = e^{\frac{-\alpha^2 u^2}{2}}$. Therefore $\mathcal{L}(X)$ is $N(0, \alpha^2)$.

Appendix B

[C] is not a vector lattice

By defining a particular operator in [C] we show that [C] is not a lattice. In section 4 it was noted that A^+ is a member of [C] iff A^+1 is continuous.

Define A by $Ag(y) = 0$ if $y \leqslant 0$

$= g(y) - g(0)$ if $y > 0$.

Then Ag is a continuous bounded function,

$A(ag + bf)(y) = aAg(y) + bAf(y)$, if $y \leqslant 0$, and

$A(ag + bf)(y)$

$= ag(y) + bf(y) - ag(0) - bf(0)$

$= aAg(y) + bAf(y)$, if $y > 0$.

We also have

$\|Ag\| \leqslant \|g - g(0)\| \leqslant 2\|g\|$, so that $\|A\| \leqslant 2$.

Therefore $A \in [C]$.

Now $A^+1(y) = \sup\{Tg(y) : 0 \leqslant g \leqslant 1,\ g \in C\}$

so $A^+1(y) = 0$ if $y \leqslant 0$.

If $y > 0$, then

$A^+1(y) = \sup\{g(y) - g(0) : 0 \leqslant g \leqslant 1,\ g \in C\}$

$\leqslant 1$ and, since there is a $g \in C$ such that

$g(x) = 0$ if $x = 0$

$= 1$ if $x \geqslant y$,

$A^{+}{}_{1}(y) \geqslant 1$

Therefore $A^{+}{}_{1}(y) = 0$ if $y \leqslant 0$

$= 1$ if $y > 0$

This shows that $A^{+} \notin [C]$.

APPENDIX C

An alternative proof of theorem 4

The proof given here depends on the representation of continuous linear functionals on C. A well known theorem in functional analysis (D. and S., p.262; Taylor, p.196), may be stated, in this special case, as follows:

if $\hat{x}$ is a continuous linear functional on C, there is a function μ, continuous on the left, of bounded variation on R such that

$\hat{x}g = \int g(y)\ d\mu(y)$ for every $g \in C$ and $\mathrm{Var}\,\mu = \|\hat{x}\|$.

$(\|\hat{x}\| = \sup_{\|g\| \leq 1} |\hat{x}g|)$

Let $A \in [C]$ commute with every degenerate pr. op. Then $\hat{x}$, defined by $\hat{x}g = Ag(0)$, is a continuous linear functional on C. Thus by the above theorem there is a function μ, continuous on the left, of bounded variation on R such that

$Ag(0) = \hat{x}g = \int g(x)\ d\mu(x)$ and $\|\hat{x}\| = \mathrm{Var}\ \mu$.

Now $Ag(y)$

$= T_y Ag(0)$

$= A(T_y g)(0)$

$$= \int T_y g(x)\ d\mu(x)$$

$$= \int g(x + y)\ d\mu(x)$$

Also $\|A\| \leqslant |Ag(y)| + \epsilon$ for some $g \in C$, with $\|g\| = 1$, and some $y \in R$.

Therefore $\|A\| \leqslant |A(T_y g)(0)| + \epsilon$

$\leqslant \|\hat{\mu}\| + \epsilon$

$= \text{Var}\ \mu + \epsilon$, so $\|A\| \leqslant \text{Var}\ \mu$.

Also $\|A\| = \sup\limits_{\|g\| = 1} \|Ag\|$

$\geqslant \sup\limits_{\|g\| = 1} |Ag(0)|$

$= \|\hat{\mu}\|$

$= \text{Var}\ \mu$.

Thus $\|A\| = \text{Var}\ \mu$.

We will now use this to show that $\|A\| = \|A^+\| + \|A^-\|$. In section 4 we showed that $\frac{A^+}{\|A^+\|}$ and $\frac{A^-}{\|A^-\|}$ are pr. op.s . Thus there are d.f.s F_+ and F_- such that $\frac{A^\pm}{\|A^\pm\|} g(y) = \int g\ (x + y)\ dF_\pm\ (x)$. Now

$$\left\| \frac{A^{\pm}}{\|A^{\pm}\|} \right\| = \frac{A^{\pm}}{\|A^{\pm}\|} 1(0)$$

$$= \int dF_{\pm}(x)$$

$$= \text{Var } F_{\pm}$$

Therefore $\text{Var } F_{\pm} = 1$.

Because they both provide a representation of A and are both continuous on the left, $\hat{\mu} = \|A^{+}\|F_{+} - \|A^{-}\|F_{-}$ and μ are equal up to an additive constant - the proof is exactly the same as 1a. Let $\mu(x) = p(x) - n(x)$ where p is the positive variation of μ and n is the negative; then p and n are non-decreasing functions. It is assumed that $\mu(-\infty) = 0$ otherwise $\mu(x) - \mu(-\infty) = p(x) - n(x)$ (Rudin, p.102). Let Var f denote the variation of f on R, then $\text{Var}\,\mu = \text{Var } p + \text{Var } n$.

Given $\epsilon > 0$, $y \in R$ and $g \in C$ such that $0 \leq g \leq 1$, there is an $f \in C$, $0 \leq f \leq g$, such that $A^{+}g(-y) \leq \epsilon + Af(-y)$. Therefore

$$\|A^{+}\| \int g(x - y)\, dF_{+}(x) \leq \epsilon + \int f(x - y)\, d\mu(x)$$

$$= \epsilon + \int f(x - y)\, dp(x) - \int f(x - y)\, dn(x)$$

$$\leqslant \epsilon + \int g(x - y)\, dp(x)$$

Now choose g so that

$$g(x) = 1 \text{ if } x \leqslant -\frac{1}{n}$$

$$= -nx \quad \text{if} \quad -\frac{1}{n} < x < 0$$

$$= 0 \quad \text{if} \quad x \geqslant 0.$$

Then

$\|A^+\|F_+(-\infty, y - \frac{1}{n}) \leqslant \epsilon + p(-\infty, y)$, so by left

continuity of F_+

$\|A^+\|F_+(-\infty, y) \leqslant p(-\infty, y)$.

Now by above,

$$\mu(-\infty, y) = p(-\infty, y) - n(-\infty, y)$$

$$= \|A^+\|F_+(-\infty, y) - \|A^-\|F_-(-\infty, y)$$ and so

$$n(-\infty, y) - \|A^-\|F_-(-\infty, y)$$

$$= p(-\infty, y) - \|A^+\|F_+(-\infty, y) \geqslant 0$$

Therefore $\|A^-\|F_-(-\infty, y) \leqslant n(-\infty, y)$ and since

p, n, F_+ and F_- are non-decreasing,

$\|A\|$

$\leqslant \|A^+\| + \|A^-\|$

$= \|A^+\|\ \mathrm{Var}\ F_+ + \|A^-\|\ \mathrm{Var}\ F_-$

$\leq$ Var p + Var n

= Var μ

= $\|A\|$.

Therefore $\|A\| = \|A^+\| + \|A^-\|$, provided A commutes with every degenerate pr. op. This does not hold in general in C; this is shown by the following example:

define T by

$$
\begin{aligned}
Tg(y) &= g(y) && \text{if } y \leq -1 \\
&= -g(-1)y && \text{if } -1 < y < 0 \\
&= -g(1)\ y && \text{if } 0 \leq y < 1 \\
&= -g(y) && \text{if } 1 \leq y
\end{aligned}
$$

Then $T \in [C]$ and $\|T\| = \|T^+\| = \|T^-\| = 1$.

REFERENCES

Birkhoff, G. Lattice Theory American Math. Soc. Revised Edition 1948

Dunford, N. and Schwartz, J.T. Linear Operators - Part I Interscience 1958

Halmos, P.R. Measure Theory van Nostrand 1961

Loéve, M. Probability Theory van Nostrand Third Edition 1963

Rao, M.M. Two Probability Theorems and an Application Indag. Math. (Amsterdam) 1961 23, p. 551 - 559

Rudin, W. Principles of Mathematical Analysis Mc Graw-Hill 1953

Taylor, A.E. Introduction to Functional Analysis Wiley 1958

Trotter, H.F. An Elementary Proof of the Central Limit Theorem Archiv der Math 1959 10 p. 226 - 234