SOME BASIC CONCEPTS IN THE THEORY OF STOCHASTIC PROCESSES AND INTRODUCTION TO MARKOV PROCESSES

R. Vasudevan
The Institute of Mathematical Sciences
Madras - 600113, India

Basic concepts in probability theory

The central role of modern probability theory in unravelling the mysteries of nature whether for explanation or prediction can be appreciated when one realises that few mathematical disciplines have contributed to such a wide spectrum of areas from physics to number theory, from biology to econometrics. Very few disciplines have gone so deep into our thinking at the conceptual level . Games of chance attracted the attention of Pascal and Fermat. The efforts of Huyghens, Laplace, Bernoulli and Quetelet and others and even earlier studies of Darwin laid the foundations of statistical methods followed later by the works of Galton, Weldon and more quantitative studies of Fisher. The gigantic work of the members of the Russian school, Tschebychev, Markov, Bernstein, Kintchine and Kolmogorov enhanced the concepts of statistical mechanics in physics. The works of Lindberg, Feller, Long, Doob and others and also the econometricians like Slutzky and Yule introduced new important tools like sequential sampling, time-series analysis etc.

One knows very well the difference between statics and dynamics. Much of economic theory and those relating to classical statistical mechanics are erected on equilibrium structure. Now they model their studies to represent closely our dynamic changing world. Thus stochastic theory in common parlance may mean a wider theory of statistics of change.

The basis of all statistical theories is probability theory governing random processes the study of which usually begins with coin tossing experiments.

According to the frequency approach of Von Mises if A is the outcome of an experiment occurring k times in n trials then the limit

$$\lim_{n\to\infty} \frac{k}{n} \sim P(A) \qquad (1)$$

is the probability for A determined by the relative frequency of A. How large n should be is not a question that can be answered with mathematical rigor.

As an illustration let us consider the distribution of N indistinguishable Bose particles in C_j cells of phase space having energy parameter E_j. Since any number of particles can be allotted to each cell the generating function of such an arrangement is $G(u) = 1/(1-u)$ and for C_j cells $G(u,C_j) = (1-u)^{-C_j}$. Hence the coefficient of the term u^{N_j} will yield $[(N_j+C_j-1)!/(N_j!(C_j-1)!)]$ which is the number of ways of distributing N_j particles. This has to be minimised along with the conditions that the total number of particles is fixed and also the total energy i.e. $\sum_j N_j = N$

and $\sum_j N_j E_j = E$. Under such conditions this yields the Bose-Einstein distribution. Derivations for Maxwell-Boltzman and Fermi-Dirac statistics can be arrived at similarly (cf. e.g. Ramakrishnan [1] for details).

In the axiomatic approach all the outcomes of an experiment constituting a set of events is described as the sample space Ω of events. Subsets of Ω, say A,B,C,.. etc. composed of elementary events $\omega \in \Omega$ may be the events we are interested in. Any field $\mathcal{B}$ on Ω is called a Borel field (or σ field) $\mathcal{B}$ on Ω if it is closed under denumerable unions and intersections of the subsets $A_1, A_2, A_3, \ldots$ etc. i.e.

$$\begin{aligned} &(1) \quad \text{if } A_1, A_2, \ldots, A_n \in \mathcal{B} \quad \text{then } \bigcup_{N=1}^{\infty} A_n \in \mathcal{B} \\ &(2) \quad \text{if } A_1, A_2, \ldots, A_n \in \mathcal{B} \quad \text{then } \bigcap_{n=1}^{\infty} A_n \in \mathcal{B} \end{aligned} \tag{2}$$

Also the compliments $\bar{A}_i$ for all i and unions and intersections of $\bar{A}_i \in \mathcal{B}$ for $i = 1,2, \ldots, n \ldots$ The members of a given Borel field are called $\mathcal{B}$ measurable. Let the collections a include all A_i's and $\bar{A}_i$'s and their unions and intersections. Then we assign a probability P(A) which is a set function for each member of the collection and thus we have a triplet $(\Omega, \mathcal{B}, P)$ defining our probability space. The P(A)'s satisfy the following:

(1) $0 \leqslant P(A)$ (2) $P(\Omega) = 1$ (3) If A_i's are mutually independent then $P(A_1 A_2 \ldots A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$ while if A_1 and A_2 are not mutually exclusive we have

$$(4) \quad P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \tag{3}$$

For illustration let us consider a single die with six faces. The space Ω is the totality of out comes $\Omega = \{1,2,3,4,5,6\}$ and the class given by $0, \Omega, \{1,3,5\}$ and $\{2,4,6\}$ may be a chosen field $\mathcal{B}$ on Ω. A number P(A) is assigned to each event in the class a of events.

If P(AB) is the joint probability of both A and B occuring and P(A|B) is the conditional probability that A occurs given that B has occurred then P(AB)=P(A|B)P(B) =P(B|A)P(A). Hence we can arrive at the generalized Bayes' theorem that for any set of $\mathcal{B} \subset \Omega$ [If (1) $\bigcup A_i = \Omega$ and (2) $A_i A_j = \emptyset$ for $i \neq j$ where $\emptyset$ is null] as:

$$P(A_j | B) = \frac{P(A_j) P(B | A_j)}{\sum_i P(A_i) P(B | A_i)} \tag{4}$$

Examples of discrete distributions are:

(a) $P(x=k) = \binom{n}{k} p^k q^{n-k}$ (Binomial distribution-k denotes the number of successes in n trials)

(b) $P(x=n) = [\exp(-\lambda)] \left(\frac{\lambda^n}{n!}\right)$ (Poisson distribution giving probability for n events) (5)

Examples of continuous distributions are

(a) $f(x) = \frac{1}{\pi(1+x^2)}$; $-\infty < x < \infty$ (Cauchy distribution)

(b) $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-(x-\mu)^2/2\sigma^2\right]$ (Gaussian distribution) (6)

There are a large number of discrete and continuous distributions used in statistical modelling of phenomena. In the so called Pearsonian systems the frequency function f(x) satisfied the equation

$$\frac{df}{dx} = \frac{x+a}{c_o+c_1x+c_2x^2} \tag{7}$$

with different values for the constants c_o, c_1 and c_2. In many applications we are interested only in the statistical features like mean and moments of the distributions. Hence we describe a few transformations of the frequency function which yeild the desired information in a simple fashion

(1) The probability generating function is

$$G(s) = \sum_n P_n s^n \text{ so that } P_n = \frac{1}{n!} \left.\frac{d^nG(s)}{ds^n}\right|_{s=0} \tag{8}$$

$\forall n = 0,1,2, \ldots$etc.

With E denoting expectation

$$\left.\frac{d^kG(s)}{ds^k}\right|_{s=1} = E\left[n(n-1)\cdots(n-k+1)\right] \tag{9}$$

are called the factorial moments which are related to the actual moments $E(n^k) = \sum_{r=1}^{k} C_k^r G^r(1)$ where C_k^r are the stirling numbers of the second kind

(2) If we put $s=e^t$ in (8) we get the moment generating function

$$G(e^t) = G_m(t) = \sum_{r=0}^{\infty} m_r \frac{t^r}{r!} \tag{10}$$

where m_r is the r-th moment

(3) For continuous distributions

$$G_m(x;t) = \int e^{tx}\, dF_X(x) \tag{11}$$

is the moment generating function for the random variable X and the central moment generating function is given by

$$C_m(t) = e^{-t\mu} G_m(x;t) = 1+\sum \bar{\mu}_\ell \frac{t^\ell}{\ell} \tag{12}$$

where

$$\bar{\mu}_\ell = E\left\{[x-\langle x\rangle]^\ell\right\}$$

(4) The characteristic function for the distribution function $F_X(x)$ is

$$\phi_X(t) = \int dx\ f(x) e^{itx} \tag{13}$$

and this is well defined since $|e^{itx}| \leqslant 1$ and ϕ_X is uniformly continuous in t with $\phi_X(0) = 1$. The characteristic function for $z = x_1+x_2$ where x_1 and x_2 are independent random variables is $\phi_z(t) = \phi_{X_1}(t)\phi_{X_2}(t)$ the product of the two characteristic functions ϕ_{X_1} and ϕ_{X_2}.

(5) The cumulant generating function K(t) is given by

$$K(t) = \log \phi_X(t) = \sum_{s=1}^{\infty} K_s \frac{(it)^s}{s!} \tag{14}$$

$\{K_s\}$ being called the cumulants which for the Poisson distribution are given by $K(t) = \lambda(e^{it}-1) = \lambda \sum_{s=1}^{\infty} \frac{(it)^s}{s!}$ and all the cumulants are the same λ the mean of the Poisson distribution.

Many limit theorems can be easily obtained using the concept of characteristic or generating functions. For example the generating function for the binomial distribution in $G_n(u) = [q+pu]^n$ where p is the success probability. If as n becomes large p is small enough so that $np = \lambda$ = a constant then $G_n(u) = [1+\frac{\lambda}{n}(u-1)]^n \xrightarrow[n\to\infty]{} \exp[\lambda(u-1)]$ which is the generating function for the Poisson distribution. Similar limiting distributions as contemplated in the central limit theorem, the law of large numbers etc. can be arrived at using characteristic functions.

If a random variable $X(\omega) = x$ corresponding to an element Ω of the sample space and if $Y = g(x) = g(X(\omega))$ is a function of the random variable and g is a mapping and $f_x(x)$ is the probability frequency function of the random variable X, the frequency function of the random variable Y is

$$f_Y(y) = \frac{f_X(x_1)}{|g'(x_1)|} + \frac{f_X(x_2)}{|g'(x_2)|} + \dots + \frac{f_X(x_n)}{|g'(x_n)|} \tag{15}$$

where $x_1, x_2 \dots$ are the zeroes of the function g(x) expressed in terms of y. It is easy to see that if x is normally distributed with mean m and variance σ and if $y = \exp(x)$ the frequency function for Y variable is called log normal distribution

$$F_Y(y) = \frac{1}{\sqrt{2\pi}\sigma y} \exp\left[-(\log y-m)^2/2\sigma^2\right] \tag{16}$$

To develop a calculus of random variables we briefly sketch three types of convergences relating to sequences of random variables

(1) Convergence in probability:

Let $X_1, X_2, \dots X_n$ be a sequence of random variables and this sequence is said to converge in probability to the random variable X if

$$\lim_{n\to\infty} \left\{P_r\ |X_n-X| \geqslant \epsilon\right\} = 0; \quad X_n \xrightarrow[n\to\infty]{P} X \tag{17}$$

Similarly one can define convergence in distribution

(2) Almost certain convergence:

The sequence of variables X_n converges to X if in the limit $n \longrightarrow \infty$, almost certainly if

$$\Pr_{\text{as } n\to\infty}(X_n = X) = 1 \quad \text{and} \quad \lim_{n\to\infty} X_n(\omega) = X(\omega) \tag{18}$$

for almost all $\omega \in \Omega$. Otherwise stated $(\Pr_{\text{as } n\to\infty}[|X_n - X| > \epsilon])$ is zero for atleast one m, m being greater than n.

(3) Mean square convergence:

The sequence X_n of the random variables is said to converge in mean square (m.s) to the random variable X if

$$E(X_n^2) < \infty \quad ; \quad E(X^2) < \infty \quad \text{and if}$$

$$\lim_{n\to\infty} E[(X_n - X)^2] = 0 \tag{19}$$

because of Chebyshev inequality

$$\Pr[|X_n - X| \geqslant \epsilon] \leqslant E[(X_n - X)^2]/\epsilon^2 \text{ for any } \epsilon > 0 \tag{20}$$

For m.s. convergence the right hand side of (20) tends to zero in the limit $n \to \infty$. Hence m.s. convergence implies convegence in probability. Almost certain convergence also implies convergence in probability.

The subject matter detailed till now is contained in the literature cited in [1-7] at the end.

Stochastic Processes

Let Ω be a set whose Borel field of subsets generates events. $\{X(\omega,t); \omega\in\Omega, t\in T\}$ is called a stochastic process which depends on two variables ω (event space) and t (the index set). The index t denotes time when the process evolves in time. X may be a real valued function for each ω with its domain on T or a random variable on the probability space (Ω, F, P) at each t. For every finite set of t-values $t_1, t_2, \ldots, t_n$ the corresponding random variables $X_1(\omega_1, t_1)$, $X_2(\omega_2, t_2)$, ... $X_n(\omega_n, t_n)$ have an n-dimensional distribution function

$$F_n(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n) = \Pr[X_n(t) \leqslant x_n, X_{n-1}(t) \leqslant x_{n-1}, \ldots, X_1(t) \leqslant x_1] \tag{21}$$

These F_n's should satisfy

(1) The symmetry condition i.e. it is symmetric in all pairs of (x_j, t_j) values which means F_n's are invariant for the same permutations of x_j and t_j.

(2) The Kolmogorov consistency condition which leads to the following

$$F_n(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n) = F_{n-1}(x_1, x_2, \ldots, x_{n-1}; t_1, t_2, \ldots, t_{n-1}) \tag{22}$$

when $x_n \longrightarrow \infty$

Since both variables x and t can be either discrete or continuous there are four categories of stochastic processes.

Stationary stochastic processes are characterised by the fact that the finite-dimensional joint distributions are invariant under translations of the origin. It is often illustrated that

$$\mathrm{Cov}\,[X(t)X(t+h)] = E[X(t)X(t+h)] - E[X(t)]\,E[X(t+h)] \quad (23)$$

is only a function of h for all $t \in T$. If X(t) process is such that

$$E\{[x(t) - E(x(t))][x(s) - E(x(s))]\} = R(t-s) \quad (24)$$

then the process is wide sense stationary or second order stationary. Then (1) if $R(\tau)$ in (24) is continuous for all τ and has second derivatives, X(t) is differentiable in the m.s. sense (2) if $R(\tau)$ is nonnegative definite for all τ, (τ being real) the process is stationary (3) Also $R(\tau)$ for a mean square continuous stationary process has the mathematical form of the characteristic function (Fourier transform) of a distribution function (Bochner's theorem) [2]. Stationarity and normality of X(t) vis a vis the nature of $R(\tau)$ will be dealt with later while discussing gaussian Markov processes.

The power spectrum of a process X(t) is given by the Fourier transform

$$F_X(\omega) = \int_{-\infty}^{\infty} X(t)\,\exp(-i\omega t)\,dt \quad (25)$$

$F_X(\omega)$ being a generalised function. If $G(\omega_1,\omega_2) = \langle F_X(\omega_1)F_X(\omega_2)\rangle$ and if it so happens that $G(\omega_1,\omega_2) = G(\omega)\,\delta(\omega_1-\omega_2)$ then $G(\omega)$ is the power spectrum of the stationary process and its different frequency components are uncorrelated.

A stochastic process defined over the interval (0,T) can be expanded in terms of a complete set of orthogonal functions $\phi_i(t)$ such that [10]

$$X(t) \underset{N\to\infty}{=} \sum_{i=1}^{N} \chi_i\phi_i(t), \qquad \int_0^T \phi_i(t)\phi_j(t)dt = \delta_{ij} \quad (26)$$

The coefficients χ_i are given by $\chi_i = \int_0^T X(t)\phi_i(t)dt$. If we want the coefficients χ_i to be uncorrelated then we should find the function ϕ_i such that

$$\int_0^T R(t,t')\phi_j(t')dt' = \mu_j\phi_j, \quad 0 \leqslant t \leqslant T \quad (27)$$

assuming that $\langle X(t)\rangle = 0$. This is the main content of the Karhunan-Levy expansion of a stochastic process X(t). Some of the interesting properties of these μ's are

$$\sum_{i=1}^{\infty} \mu_i = \int_0^T G_x(t,t)\,dt; \quad \sum_{i=1}^{\infty} \mu_i^2 = \iint |G_x(t_1,t_2)|^2 dt_1dt_2$$

$$G_x(t_1,t_2) = \sum_{i=1}^{\infty} \mu_i\phi_i(t_1)\phi_i(t_2) \quad \text{etc.} \quad (28)$$

and if X(t) is stationary then in the limit of large T, much larger than the width of $G_x(\tau)$ we find that

$$\phi_n \sim \frac{1}{\sqrt{T}} \exp(i\omega_n t) \quad \text{where} \quad \omega_n = \frac{2\pi n}{T} \tag{29}$$

Point processes [1,8-12]

A process whose realisations are related to a series of point events occuring in a continuous one-dimensional parameter space (such as time etc.) are point processes which are distinguished by their times of occurrence. The Poisson events and the renewal processes have been studied even before a theory of point process was developed. Studies of assemblage of particles distributed in phase space are also a type of point process. A powerful tool for the study of point processes is the product density technique of Ramakrishnan [2,12]. The central quantity of interest is $dN(x,t)$ denoting the number of entities with parametric values between x and x+dx at time t. This is proportional to dx. It is assumed that the probability that there is more than one particle in that range is of order $O(dx)$ and hence the probability that there are n particles in that range is given by

$$P(1) = f_1(x,t)dx, \quad P(n) = 0, \quad n>1, \quad P(0) = (1-f_1(x,t)dx) \tag{30}$$

Hence $E[n^m] = E[dN(x,t)^m] = \sum P(n)n^m$

$$= E[dN(x,t)] + O(dx) = f_1(x,t)\,dx \tag{31}$$

If $dN(x,t)$ takes the value unity if the variable x happens to lie between x and x+dx then $E[dN(x,t)]$ is the probability that an entity with parametric values between x and x+dx occurs at t. The regularity condition is that $\sum_{n\geq 2} P(n,x;x+\Delta) = O(\Delta)$. Though ideas of this type for first order densities existed in earlier works [13,14] this concept has been deeply analysed and perfected as a sophisticated tool for the study of point processes by Ramakrishnan [12] and many applications and development of these ideas have been made by Ramakrishnan and his group involving higher order correlation functions [1,11].

$f_1(x,t)dx$ is called the first order product density and the expectation of the number in a given interval $[a,b]$ is

$$E[N(a,t)-N(b,t)] = \int_a^b f_1(x,t)\,dx \tag{32}$$

The second order product density and higher order correlations are given by

$$E[dN(x_1,t)dN(x_2,t)] = f_2(x_1,x_2,t)\,dx_1dx_2 \quad \text{provided } x_1 \text{ and } x_2 \text{ do not overlap}$$

$$E[dN(x_1,t)dN(x_2,t)\ldots dN(x_n,t)] = f_n(x_1,x_2,\ldots,x_n,t)dx_1dx_2\ldots dx_N \tag{33}$$

For the second moment of the number of points we find

$$E\left\{[N(a,t) - N(b,t)]^2\right\} = \int_a^b\int E[dN(x_1,t)dN(x_2,t)] \tag{34}$$

and due to the singular behaviour if x_1 and x_2 coalesce the integral in the equation splits into two regions to yield

$$E\left\{[N(a,t)-N(b,t)]^2\right\} = \int_a^b f_1(x,t)\,dx + \int_a^b\int_a^b f_2(x_1,x_2\ t)\,dx_1 dx_2 \tag{35}$$

Similarly any r-th moment of the number of entities in a given region of x-space is given by

$$E\left\{[N(a,t)-N(b,t)]^r\right\} = \sum_{s=1}^{r} C_s^r \int_{x_1}\int_{x_2}\cdots\int_{x_s} f_s(x_1,x_2\ldots x_s;t)\,dx_1 dx_2\ldots dx_s \tag{36}$$

where C_s^r denotes the number of (r-s) fold degeneracy of an r-fold product and $\left[C_s^r\right]$ are the Stirling numbers of the second kind. So the moments are related to the factorial moments as

$$E[N^r] = \sum_{s=1}^{r} C_s^r \langle N(N-1)\ \ldots.(N-s+1)\rangle \tag{37}$$

The product densities are related to what are called the Janossy densities $J_n(x_1,x_2,\ldots;t)$ which represent the probability that there are exactly n particles distributed in the parametric space at x_i's, $i = 1,2,\ldots,n$. Since particles are indistinguishable the probability for the occurence of n particles is

$$P(n,t) = \frac{1}{n!}\int_\Omega J_n(x_1,x_2\ldots,x_n;t)\ dx_1 dx_2\ldots dx_n \tag{38}$$

and the product density of order h which expresses the probability of finding particles or entities in the ranges (x_1,x_1+dx_1), (x_2,x_2+dx_2)....etc. irrespective of what happens in other ranges is given by

$$f_h(x_1,x_2,\ldots x_h;t) = \sum_{n=h}^{\infty} \frac{1}{(n-h)!}\iint\cdots\int J_n(x_1,x_2,\ldots x_h,x_{h+1},\ldots x_n;t)\ dx_{h+1}dx_{h+2}\ldots dx_n \tag{39}$$

It follows from (38) and (39) that [15,16]

$$P(n,t) = \frac{1}{n!}\sum_{k=n}^{\infty}\frac{(-1)^{k-n}}{(k-n)!}\iint\cdots\int f_k(x_1,x_2\ldots,x_k;t)\,dx_1\ldots dx_k \tag{40}$$

The moment generating function $\theta(u) = \sum_{n=0}^{\infty} P_n \exp(inu)$ in view of the above relations can be expressed in terms of the product density functions as follows

$$\theta(u;t) = \sum_{s=0}^{\infty}\frac{(e^{iu}-1)^s}{s!}\iint_\Omega\cdots\int f_s(x_1,x_2,\ldots x_s;t)\,dx_1 dx_2\ldots dx_s \tag{41}$$

Since moments can be obtained from $\theta(u;t)$ we easily see from (36) that the C_s^r coefficients are given by

$$C_s^r = \left\{\frac{1}{i^r}\frac{\partial^r}{\partial u^r}\left[\frac{(e^{iu}-1)^s}{s!}\right]\right\}_{u=0} = \frac{1}{s!}\sum_{k=0}^{s} k^r\binom{s}{k}\ (-1)^{s-k} \tag{42}$$

The product densities have separable parts and nonseparable cluster terms called irreducible or connected parts as given below

$$f_1(x;t) = g_1(x;t) \; ; \; f_2(x_1,x_2) = g_1(x_1)g_1(x_2) + g_2(x_1,x_2) \; ;$$

$$f_3(x_1,x_2,x_3) = g_1(x_1)g_1(x_2)g_3(x_3) + 3\{g_1g_2\} + g_3(x_1,x_2,x_3) \tag{43}$$

Herein we have suppressed the t dependence and the symbol $\{ \quad \}$ means proper symmetrisation. Similar expression for higher order correlation functions can be explicitly written down.

The product density generating functional can be written down as

$$D_\Omega(v) = 1+ \sum_{s=1}^{\infty} \frac{1}{s!} \iint_\Omega \cdots \int f_s(x_1,x_2,\ldots,x_s)v(x_1)v(x_2)\ldots v(x_s) \quad dx_1dx_2\ldots dx_s \tag{44}$$

and f_s can be obtained as the sth order functional derivative of D_Ω with respect to $v(x_1),v(x_2)\ldots.v(x_s)$ and it can also be shown that

$$D_\Omega(v) = \exp\left\{ \sum_{s=1}^{\infty} \frac{1}{s!} \iint_\Omega \cdots \int g_s(x_1,x_2,\ldots x_s)dx_1dx_2\ldots dx_s \right\} \tag{45}$$

where $g_s(x_1,x_2,\ldots x_s)$ is the sth order cluster function. The probability generating function $G(z) = \sum z^n P(n,t) = D_\Omega(z-1)$ and hence for the Furry distribution with $f_n(x_1,x_2,\ldots x_n) = n!f_1(x_1)f_1(x_2)\ldots.f_1(x_n)$ we have

$$G(z) = (1-z \int_\Omega f_1(x_1)dx_1)^{-1} \tag{46}$$

If a light beam incident on a photodector with sensitivity α ejects electrons proportional to the intensity $I(t)$ of the beam the product density of getting an electron in the time interval $(t,t+dt)$ is given by $f_1(t)dt = \langle \alpha I(t) \rangle dt$ where $\langle I(t) \rangle = \langle V(t)V^*(t) \rangle$ V being the analytic signal characteristic of the light which has a gaussian distribution for chaotic light. The product density generating functional

$$D_\Omega(v) = \left\langle e^{\alpha \int_0^T I(t)u(t)dt} \right\rangle \tag{47}$$

for this case. This, by (45), is given in terms of the connected parts as

$$D_\Omega(u) = \exp\left\{ \sum_{s=1}^{\infty} \frac{\alpha^s}{s!} \int_0^T\int \cdots \int \langle I(t_1)I(t_2)\ldots I(t_s) \rangle_c \, dt_1dt_2\ldots dt_s \right\} \tag{48}$$

The cluster part is given by

$$\alpha^s \langle I(t_1)I(t_2)\ldots I(t_s) \rangle_c = \Gamma^s(t_1,t_2,\ldots t_s) \tag{49}$$

where Γ^s is the s-th order coherence function. We have seen that the probability generating function $G(z)$ can be obtained by simply getting $D_\Omega(z-1) = G(z)$ and hence

$$G(z) = \exp \sum_{s=1}^{\infty} \frac{(z-1)^s}{s!} \iint_0^T \cdots \int \Gamma^s(t_1,t_2,\ldots,t_s)dt_1dt_2\ldots dt_s \tag{50}$$

$\Gamma^s(t_1,t_2,\ldots t_s)$ can be taken as $(s-1)!(\alpha IT)^s$ for the case of chaotic light, I being a constant. Then we have [11,17,18]

$$G(z) = (1+\alpha IT-\alpha zT)^{-1} \tag{51}$$

Hence inverting we get

$$P(n,T) = \frac{1}{(1+\bar{n})} \frac{1}{(1+\frac{1}{\bar{n}})^{n}} \tag{52}$$

which is a Boson distribution. For mixing of different types of light beams etc., the reader is referred to [17,18] .

Multiple product densities [19] where a twin or an i-tuple can constitute a point process with parametric values between x and x+dx can be formulated. In this case $f_1^i(x) = E[dN^i(x)]$ and $f_2^{i_1,i_2}(x_1,x_2) = E\,dN^{i_1}[(x_1)dN^{i_2}(x_2)]$ where f_1^i is the product density of first order for the occurrence of an i-tuple between x and x+dx and the mixed density $f_2^{i_1,i_2}(x_1,x_2)$ denotes the probability that there is an i_1-tuple in (x_1,x_1+dx_1) and an i_2-tuple in (x_2,x_2+dx_2). These are useful in the birth of pions etc., and in neutron multiplication problems [10,19,23].

A sequent product density such as $F_2(E_1,E_2;t_1,t_2|E_o,o)dE_1,dE_2$ represents the probability of occurrence of particles in the energy ranges (E_1,E_1+dE_1) and (E_2,E_2+dE_2) at times t_1 and t_2 respectively, given that there was a particle at t=0 with energy E_o. These find great use in cosmic ray shower theories. When $t_1 < t_2$

$$F_2(E_1,E_2;t_1,t_2|E_o) = \int f_2(E_1,E_2'|E_o,t_1)\; f_1(E_2|E_2';t_2-t_1)\; dE_2'$$

$$+ f_1(E_1|E_o;t_1)\; f_1(E_2|E_1,t_2-t_1) \tag{53}$$

and when $t_1 \rightarrow t_2$ the limiting case is represented as

$$F_2(E_1,E_2|t_2,t_2) = f_2(E_1,E_2;t_2) + f_1(E_1,t_2)\;\delta(E_2-E_1) \tag{54}$$

These were formulated in [20] and evolutionary product densities in product space Ω of both x and t have been widely used in cosmic ray theories [21].

Starting with probability generating function $F(u) = \sum_{n=o}^{\infty} u^n P(n)$ we can express

$$F(u) = \exp\left\{\sum_k (u^k-1)c_k\right\} \tag{55}$$

where c_k's are called 'combinants'. They are a measure of the deviation of the probability function P(n) from a Poisson. Assuming that $P(0) \neq 0$ for the Poisson case $F(u) = \exp[\bar{n}(u-1)]$. Hence only one combinant $c_1 = \bar{n}$ exists in this case. For a general correlated process [22]

$$\log F(u) - \log P(0) = \sum_{k=1}^{\infty} c_k u^k,\; P(0) = \exp\left\{-\sum_{k=1}^{\infty} c_k\right\} \tag{56}$$

c_k's are expressible in terms of Pr's upto the same order r = k and conversely if c's are known upto any order probabilities upto that order can be found. Also it can be shown by using this concept and the relations in equations (59,56 etc.) that the cumulants of a probability distribution can be arrived at as

$$K_r = \sum_s c_s^r \tau_s \tag{57}$$

where τ_s are integrals over cluster functions

$$\tau_s = \iint \cdots \int g_s(x_1, x_2, \ldots, x_s)\, dx_1 dx_2 \ldots dx_s \tag{58}$$

Also we can show that

$$K_r = \sum_{\ell=1}^{\infty} \ell^r c_\ell \tag{59}$$

where c_ℓ's are the combinants. For more details see references [22,(a),(b)].

Markov Processes[1,23-27]

Let a physical system occupy one of the finite and discrete set of states $\{X_n\}$, n = 0,1,2,...etc. (which are random variables) at discrete set of times $\{n\Delta\}$. The sequence of $\{X_n\}$ constitute a Markov chain if the random variables X_{n_m} have a conditional probability which depend only on the value of the random variable at one step earlier only i.e.

$$Pr\left\{X_{n_m} = x_{n_m} \mid X_{n_{m-1}} = x_{n_{m-1}}, X_{n_{m-2}} = x_{n_{m-2}}, \ldots, X_{n_o} = x_{n_o}\right\}$$

$$= Pr\left\{X_{n_m} = x_{n_m} \mid X_{n_{m-1}} = x_{n_{m-1}}\right\} \tag{60}$$

This is a discrete one step Markov chain. Simplifying the notation let us consider $P_i(m)$ the probability that the system is in the state i at the time step m and formulating the one step transition probability by

$$P_{ij} = Pr\left\{X_{m+1} = j \mid X_m = i\right\} = P_{ij}(1) \tag{61}$$

we can see that $P_{ij}(m)$, the transition probability for m steps can be found as

$$P_{ij}(m) = P_{ij}^m \tag{62}$$

Hence for a homogeneous Markov chain $\{X_n\}$ the (n+m) step transition probability satisfies the Chapman-Kolmogarov relation

$$P_{ij}(n+m) = \sum_k P_{ik}(n)\, P_{kj}(m) \tag{63}$$

Also $\sum_j P_{ij} = 1$. Such matrices with elements $[P_{ij}]$ summing upto unity along the rows (or columns) are called stochastic matrices.

A state is called a recurrent or persistent state if starting from it the ultimate return to it is a certainty. The time of first return to that state is a random variable called recurrence time. A state is called positive recurrent or null recurrent according as the mean recurrence time is finite or infinite. A periodic recurrent state is one which can be reached only at $t^{th}, 2t^{th}, 3t^{th}$....time steps.

A recurrent, nonnull, aperiodic state is called an ergodic state.

If we take an initial unconditional probability vector $\vec{\Pi}(0)$ for a finite state system and obtain the probability vector $\vec{\Pi}(n)$ at the n-th time step then we would have

$$\vec{\Pi}(n) = (\bar{P})^n \vec{\Pi}(0) \tag{64}$$

where $\bar{P}_{ij} = P_{ji}$ i.e. $\bar{P}$ is the transpose of the transition probability matrix $[P_{ij}]$ described above. If $\vec{\Pi}(n) \underset{n\to\infty}{\longrightarrow} \vec{\Pi}$ independent of n then the chain is ergodic. Hence an irreducible and aperiodic Markov chain is ergodic if we can find a non-null solution of the equation

$$\vec{X} = \bar{P}\vec{X} \text{ for } \sum_i |X_i| < \infty \tag{65}$$

Markov processes in continuous time like the Poisson process $P(n,t)$ or the birth-death processes represent evolution processes without memory. Given the state of the system at time t the development of the system between t and $t+\Delta$ depends only on the state of the system at time t and independent of the states prior to t. This is the probabilistic analogue of the deterministic trajectories in mechanics. An elementary one dimensional random walk where the state at n-th step $Y(t_n) = \sum_{i=1} X_i$ with X_i's as independent random variables corresponding to the sizes of the steps $\{i\}$, is a Markov process.

Let us take discrete states and continuous time and consider the transition probability, $\Pi(X_i|X_k,t)$, that the system goes to the state X_i from the state X_k in the time interval $(0,t)$. Then by the Chapman-Kolmogorov relation

$$\Pi(X_i|X_k;t+\Delta) = \sum_j \Pi(X_j|X_k;t)\,\Pi(X_i|X_j;\Delta) \tag{66}$$

Let us further assume that $\lim_{\Delta\to 0} \Pi(X_i\ X_j;\Delta) = R_{ij}\Delta$. Hence taking account of the probabilities that can exhaust the possibilities that can happen in the interval Δ we have

$$\Pi(X_i|X_k;t+\Delta) = (1-\sum_{j\neq i} R_{ji}\Delta)\,\Pi(X_i|X_k;t) + \sum_j \Pi(X_j|X_k;t)R_{ij}\Delta \tag{67}$$

In the limit $\Delta \longrightarrow 0$, this leads us to the forward differential equation

$$\frac{\partial \Pi(X_i|X_k;t)}{\partial t} = -\sum_{j\neq i} R_{ji}\,\Pi(X_i|X_k;t) + \sum_{j\neq i} R_{ij}\,\Pi(X_j|X_k;t) \tag{68}$$

By relating $\Pi(X_i|X_j;t)$ to $\Pi(X_i|X_k;t-\Delta)$ taking into consideration to possibilities in the first interval $(0,\Delta)$ the backward differential equation can be obtained as

$$\frac{\partial \Pi(X_i|X_k;t)}{dt} = -\Pi(X_i|X_k;t)\sum_{j\neq k} R_{jk} + \sum_{j\neq k} \Pi(X_i|X_j;t)R_{jk} \tag{69}$$

It is an interesting tool to derive equations governing the evolution of the moment generating function for a Markov continuous process. This can be done in a genera-

generalized fashion. The moment generating function at time $t+\Delta t$ for the $X(t)$ process is

$$M(\theta,t+\Delta t) = \underset{t+\Delta t}{E} \left\{ \exp\{(\theta X(t+\Delta t))\} \right\} \quad (70)$$

where $E_{t+\Delta t}$ is the expectation with respect to the random variable $X(t+\Delta t)$. Let us call E_t, the expectation with respect to the variable $X(t)$, and $E_{\Delta t/t}$ the expectation of the variable $X(t)$ at the end of time Δt due to fluctuations in ΔX given that at time t the variable has a value $X(t)$. Then

$$M(\theta,t+\Delta t) = \underset{t}{E}\left\{ \exp(\theta X(t)\underset{\Delta t/t}{E} \left[\exp(\theta \Delta X(t))\right]\right\} \quad (71)$$

Hence

$$\lim_{\Delta t\to o} \frac{M(\theta,t+\Delta t)-M(\theta,t)}{\Delta t} = \underset{t}{E}\left\{ \exp(\theta X(t)) \lim_{\Delta t\to o} \underset{\Delta t/t}{E} \left[\exp\ (\theta\Delta X(t))-1\right]\right\} \quad (72)$$

and hence

$$\frac{\partial M}{\partial t} = \underset{t}{E}\left\{\exp(\theta X(t))\ \psi(\theta,t,X)\right\} = \psi(\theta,t,\frac{\partial}{\partial\theta})M(\theta,t) \quad (73)$$

where $\psi(\theta,t,X) = \underset{\Delta t/t}{E} \left\{\exp\ (\theta\Delta X(t))-1\right\}$

For a finite state process in continuous time let

$$\Pr\left\{\Delta X(t) = j|X(t)\right\} = f_j(X)\,\Delta t;\ j \neq o$$

$$\Pr\left\{\Delta X(t) = 0|X(t)\right\} = 1-\sum_j f_j(X)\,\Delta t \quad (74)$$

Hence $\psi(\theta,t,X) = \sum_{j\neq o} f_j(X)\left[\exp\ (\theta_j)-1\right]$. Thus for a pure birth and death process with $f_{+1} = \lambda X$ and $f_{-1} = \mu X$ we have the partial differential equations for M as

$$\frac{\partial M(\theta,t)}{\partial t} = \left[\lambda(\exp[\theta] -1) + \mu(\exp[-\theta]-1)\right] \frac{\partial M(\theta,t)}{\partial\theta} \quad (75)$$

with $M(\theta,o) = \exp\ (a\theta)$ if $X(0) = a$. This equation can be arrived at also by the usual Chapman-Kolmogorov techniques.

If we take continuous set of states and continuous time

$$P(y_3|y_1;t+\tau) = \int_\Omega P(y_3|y_2;\tau)\cdot P(y_2|y_1;t)dy_2 \quad (76)$$

If for small τ we take $P(y_2|y_1;\tau) = (1-a_o(y_1)\tau)\ \delta(y_2-y_1)+\tau W(y_2|y_1)$ where $a_o(y_1) = \int_\Omega W(y_2|y_1)dy_2$ we have the socalled Master equation which after suppressing the intial state symbol y runs as

$$\frac{\partial P(y,t)}{\partial t} = \int_\Omega\left[W(y|y')P(y';t) -W(y'|y)P(y;t)\right] dy' \quad (77)$$

The solution of these types of equations have been dealt with in [26].

Product densities and electromagnetic cascades:

There are two important processes which are key factors in the development of the shower of electrons and photons: (1) Bremstrahlung process in which an electron of energy E_o emits a photon of energy (E_o-E') and proceeds with energy E' in its passage through a medium and the cross section for such a process is $R_1(E'|E_o)dE'$ per unit thickness; (2) Pair production process in which a photon of energy E_o produces an electron of energy E' and a positron with the rest of the energy. The cross section or probability for such a process is $R_2(E'|E_o)\ dE'$ per unit depth. With an initial particle we use the method of invariant imbedding [29] to obtain a complete set of equations for the product densities for finding an electron or a photon of given energy at a given depth. The philosophy of this method is to study what happens at the initial depth Δ and then to express the process by the same function for the rest of the depth $t-\Delta$ with new initial conditions. Thus we imbed the original problem in a class of similar problems leading to functional equations.

To this end we define first order densities $\{f_1^i(E,t|E_o)\}$ which denote the probability that there is an electron with energy in the range $(E,E+dE)$ at a depth t, the shower being initiated by a particle i with energy E_o. If $i = 1$ it is an electron initiated shower and $i = 2$ means a photon initiated shower. The imbedding equations are

$$\frac{\partial f_1^i(E,t|E_o)}{\partial t} = - f_1^1(E,t|E_o)\int_0^{E_o} R_1(E'|E_o)dE' + \int_0^{E_o} R_1(E'|E_o)\left[f_1^1(E,t|E') + f_1^2(E,t|E_o-E')\right] dE'$$

$$\frac{\partial f_1^2(E,t|E_o)}{\partial t} = 2\int_0^{E_o} R_2(E'|E_o)\ f_1^1(E,t|E')dE' - f_1^2(E,t|E_o)\int_0^{E_o} R_2(E'|E_o)dE' \qquad (78)$$

With proper initial conditions these equations can be solved using Mellins transforms. We can also formulate equations for higher order product densities. In every order one gets after taking the transforms, an equation of the type

$$\frac{d}{dt}\vec{F}_n = [A]\ \vec{F}_n + \vec{\Psi} \qquad (79)$$

By this imbedding approach one always gets A to be a matrix of dimensions 2 X 2 and $\vec{\Psi}$ is also two component vector. Thus the dimensionality of the problem is very much reduced [30]. In the usual methods [28a-d] A will be a matrix of order $2^n x 2^n$, etc. Also there are no mixed densities as those occuring in earlier methods. These equations describe the evolution of the correlations at $t+\Delta$ in terms of the spectrum at time t and do not yield higher correlations at later times. In this sense these equations, though of the same type as the Chapman-Kolmogorov equations, can be called sub-markovian. Many more processes of this type of branching processes can be treated by imbedding methods [28b,1].

Fokker-Planck equations[1,23,24,26,31]

Let us start with Chapman-Kolmogorov equations for the continuous process in continuous time. Then

$$P(x|y;t+\Delta) = \int_{-\infty}^{\infty} P(z\, y;t)\; P(x|z,\Delta)dz \tag{80}$$

and calling $\theta(u,z) = \int \exp[iu(x-z)]\; P(x|z;\Delta)dx$ we can rewrite (80) as

$$P(x|y;t+\Delta) = \frac{1}{2\pi} \iint \exp[-iu(x-z)]\; \theta(u,z)\; P(z|y;t)dzdu \tag{81}$$

We call the conditional moments $m_s = E\,[(x-z)^s]_{P(x|z;\Delta)}$ and note that

$\frac{1}{2\pi}\int \exp[-iu(x-z)]du(iu)^s = \left(-\frac{\partial}{\partial x}\right)^s \delta(x-z)$. We can expand the r.h.s of (81) in terms of m_s since $\theta(u,z) = 1+\sum_{s=1}^{\infty} \frac{(iu)^s}{s!} m_s$. Then we find that

$$P(x|y;t+\Delta) = P(x|y;t) + \sum_{s=1}^{\infty} \left(-\frac{\partial}{\partial x}\right)^s \frac{1}{s!}[m_s(x)P(x|y;t)] \tag{82}$$

If $m_s = k_s\Delta$ for each s, we obtain as $\Delta \longrightarrow o$ the following generalised Fokker-Planck equation

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial x}[k_1(x)P(x|y;t)] + \frac{1}{2!}\frac{\partial^2}{\partial x^2}[k_2(x)P(x|y;t)] + \frac{1}{3!}\frac{\partial^3(k_3P)}{\partial x^3} + \dots \tag{83}$$

For white noise, jump moments higher than k_2 are zero. Then if $k_1 = \beta x$ and $k_2 = 2D$ we get the equation

$$\frac{\partial P}{\partial t} = \beta \frac{\partial}{\partial x}(xP) + D\frac{\partial^2 P}{\partial x^2} \tag{84}$$

the solution of which is the Uhlenbech-Ornstein process with

$$P(x|y;t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-(x-\bar{y})^2/2\sigma^2] \tag{85}$$

with $\bar{y} = y\exp(-\beta t)$ and $\sigma^2 = (D/\beta)(1-\exp[-2\beta t])$. Instead of (83) we can also get the backward Fokker-Planck differential equation

$$\frac{\partial P(x\, y;t)}{\partial t} = K_1(y)\frac{\partial P(x|y;t)}{dy} + \frac{1}{2}k_2(y)\frac{\partial^2 P(x|y;t)}{\partial y^2} \tag{86}$$

If the Fokker-Planck equation has boundaries a and b and if $b = -\infty$, we can define the first passage time T of the process to reach the boundary a as [27(a)]

$$T = \inf\{t \mid x(t) > a\} \tag{87}$$

The distribution function for the first passage time $T(X_o)$ is

$$G(x_o;a;t) = Pr\{T(x_o) \leqslant t\} \tag{88}$$

and $\frac{dG}{dt} = g(x_o;a;t)$ the density function for the first passage time to be in $(t,t+dt)$. Let $f(x_o;x;t)$ be the unbounded solution of the Fokker-Planck equations, which is the density function for the particle to lie in the interval $(x,x+dx)$ at time in (t,t+dt) starting from initial value x_o. If $f^*(x_o;x,\theta)$ and $g^*(x_o;a;\theta)$ are Laplace transforms with respect to time t and if $x_o < a < x$ we have [16,23].

$$f(x_o;x;t) = \int_0^t g(x_o;a;\tau)\, f(a;x;t-\tau)\,d\tau \tag{89}$$

and

$$g^*(x_o;a;\theta) = f^*(x_o;x;\theta)/f^*(a;x;\theta) \tag{90}$$

First passage times for jump processes can be found for special types of exponential jumps by introducing compensation functions (to take care of the boundaries) in the equation for the transition probability $f(x;t|x_o)$ and treating the equation as free boundary problem. These lead to interesting results for the first passage densities in terms of the free Green's functions. These methods can also be applied to moving boundaries (For details cf.[32-34/27].

References

1. A. Ramakrishnan, Handbuch der Physik III/2, 413 (Springer-Verlag, 1959).
2. M.S. Bartlett, "An introduction to stochastic processes" Cambridge University Press, (1955).
3. A. Papoulis, "Probability, random variables and stochastic processes" McGraw Hill, New York.
4. Yu.A. Prophorov and Yu.A. Rozanov, "Probability Theory" Springer-Verlag (1969).
5. E. Lukas, "Characteristic functions", Griffin Monographs No.5, London (1960).
6. S.K. Srinivasan and K.M. Mehta, "Probability and random processes" end Edn. Tata McGraw Hill, New Delhi, (1981).
7. W. Feller, "An introduction to probability theory and applications" Vols. I and II, John Wiley, New York (1972).
8. N.J. Bailey, "The elements of stochastic processes" John Wiley, New York, (1964).
9. D.L. Snyder, "Random point processes", Wiley Interscience, New York (1975).
10. B. Saleh, "Photoelectron statistics" Springer Series in Optical Sciences, (1978).
11. S.K. Srinivasan, "Stochastic point processes", Griffin statistical monographs, London (1974).
12. A. Ramakrishnan, Proc. Camb. Phil. Soc. 46, (1950) 595.
13. D.G. Kendall, J. Roy. Stat. Soc. B11 (1949) 230.
14. J. Yuon, "Fluctuations en densite actualites scientifique et industriells" Paris.
15. P.L. Kuznetzov, R.L. Stratanovitch and V.I. Tikhonov, "Nonlinear transformations of stochastic processes" Pergamon press, London (1965) Vol.I.
16. R.L. Stratonovich "Topics in the theory of random noise" Gorden Breach, New York (1963).
17. R. Vasudevan and S.K. Srinivasan, Nuovo Cimento Ser. V 47 (1967) 183.
18. R. Vasudevan and S.K. Srinivasan, Nuovo Cimento Ser. B 8 (1972) 278.
19. A. Ramakrishnan and S.K. Srinivasan, Bull. Math. Biophys. 20 (1958) 288.
20. A. Ramakrishnan and T.K. Radha, Proc. Camb. Phil. Soc. 57 (1961) 843.
21. S.K. Srinivasan, "Stochastic theory and Cascade processes", Elsivier, New York (1969).

22a.R. Vasudevan, P.R. Vittal and K.V. Parthasarathy (to be published) Matscience Preprint.

22b.Kauffman S.K. and Gyulassy M., J. Phy A (1979) 11 p.p 715

23. A.T. Barucha Reid, "Elements of the theory of Markov processes" McGraw Hill, New York (1960).
24. Nelson Wax, "Selected papers on noise and stochastic processes" Dover, New York (1956).
25. N.V. Prabhu, "Stochastic processes", John Wiley, New York (1970).
26. N.V. van Kampen, "Stochastic processes in physics and chemistry" North Holland (1981).

27a. D.R. Cox and H.D. Miller, "Theory of stochastic processes" 2nd edn., Metheun, London (1965).
27b. T.E. Harris, "Theory of branching processes", Springer-Verlag (1963).
28a. S.K. Srinivasan, "Stochastic theory and Cascade processes" Elsivier, New York (1969).
b. A. Ramakrishnan, "Elementary particles and Cosmic rays" Pergamon (1962).
c. A. Ramakrishnan and S.K. Srinivasan, "Proc. Ind. Acad. Sci. A 44 (1956) 26.
d. A. RAmakrishnan and P.M. Mathres, Prog. Theor. Phys. 11 (1954) 95.
29. R.E. Bellman and M. Wing, "Introduction to the theory of invariant imbedding", Acad. Press (1976).
30. R. Vasudevan, A. RAmakrishnan and S.K. Srinivasan, J. Math. Anal. Appl. 11 (1965) 278.
31. R. Vasudevan and S.K. Srinivasan, "An introduction to the theory of random differential equations" Elsivier, New York (1971).
32. J. Kielson, "Green's function methods in probability theory" Griffin Monograph London (1965).
33. R. Vasudevan and P.R. Vittal, J. Theor. Neurobiology 1 (1982) 219.
34. R. Vasudevan and P.R. Vittal, J. Neurological Research Vol.4 (1982) 1/2, p.63-87.