UNITED ARAB REPUBLIC

THE INSTITUTE OF NATIONAL PLANNING

التخطيط القومى

Memo No 556

Notes On Statistical Methods

By

Dr. Abd El-Azim Anis

Operations Research Center

March 1965.

Preface

This is a short-period course on statistical methods designed for mathematicians and engineers whose work needs a thorough knowledge of statistical methods.

Because of considerations of time. It was necessary to concentrate on the theory of probability and distributions, leaving the applied side to another course given by Dr. M. W. Mahmoud.

By using the theory of sets, and the matrix notation, it is hoped that the course would make a more sound approach to the theory of probability, and give shorter proofs to a good number of theoriems. It is hoped also that this would help in the field of applications.

The course also includes an introduction to stochastic processes and Random - Walk problems which should be useful to those interested in applications in this field, among engineers and research students in economics, biology and other related fields.

I take this opportunity to thank Mrs. Mary Naguib for the generous help she gave in preparing this course, correcting the proofs, and organizing the publication of this memorandum.

A. A. Anis

9/3/1965.

1. <u>SETS</u> The possible outcomes of an experiment are called (random) <u>events</u>, which may be simple or compound, the latter being aggregates of simple events. It is convenient to represent simple events as points in a space of appropriate dimension, called the <u>observation space</u>. Compound events are then represented by sets of points.

2. Notation:

$x \in A$ means	x is a member of the set A
$\{x:K(x)\}$	the set of all x's having the property K(x)
$A = B$	the sets A, B consist of the same elements- ie.if $x \in A$ then $x \in B$ and conversely
$A \subset B$ or $B \supset A$	A is a subset of B; If $x \in A$ then $x \in B$
Union : $A \cup B$,	$\{x: x \in$ at least one of the sets $A, B\}$
Intersection: AB	$\{x: x \in A$ and also $x \in B\}$
Complement: $\bar{A}$	If S is the whole space, the complement of A (with respect to S) is $\bar{A} = \{x: x \in S$ and $x \notin A\}$
Empty set: 0	denotes the set which has no members

Disjoint sets: A and B are disjoint if $AB=0$

3. The following properties hold

Commutative laws	:	$A \cup B = B \cup A$, $AB=BA$
Associative laws	:	$(A \cup B) \cup C = A \cup (B \cup C) = A \cup B \cup C$
		$(AB)C = A(BC) = ABC$
Distributive laws	:	$A(B \cup C) = AB \cup AC$, $A \cup (BC) = (A \cup B)(A \cup C)$
Idempotence	:	$A \cup A = AA = A$
Zero and unit	:	$A \cup 0 = A$, $A0 = 0$, $AS=A$ whenever $S \supset A$
Complementation	:	$\overline{A \cup B} = \bar{A}\,\bar{B}$, $\bar{A} \cup \bar{B} = \overline{AB}$

4. Logical dictionary.

Using the representation of (1), we have the following correspondence

the set A ... the event A	Complement $\bar{A}$ ... Negation, not - A
$A \cup B$ ... disjunction, A or B	AB ... conjunction, both A & B
$A \subseteq B$ ... implication, A implies B	AB = 0 ... A,B mutually exclusive

(2) AXIOMS OF PROBABILITY (Kolmogorov)

We have a basic set E (corresponding to the observation space) whose members are the simple events. $\mathfrak{F}$ is a set of subsets of E. Then

1. $\mathfrak{F}$ is a field of sets
2. $\mathfrak{F} \supset E$
3. To each set A of $\mathfrak{F}$ is assigned a non-negative real number P(A), the prob. of A.
4. P(E) = 1
5. If AB=0, $P(A \cup B) = P(A) + P(B)$
6. If $A_1 \supset A_2 \supset A_3 \ldots \supset A_n \ldots$ and $A_1 A_2 A_3 \ldots = 0$ then
$$\lim_{n\to\infty} P(A_n) = 0$$
(where "complete additivity" if the A_i are disjoint, $P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$)

2. Basic probability laws.
(i) $P(0)=0$, $0 \le P(A) \le 1$, (ii) $P(\bar{A}) = 1 - P(A)$,
(ii) $P(A \cup B) = P(A) + P(B) - P(AB)$
(iv) $P(A \cup B \cup C) = \sum P(A) - \sum P(AB) + P(ABC)$; (v) $P(\cup A_i) \le \sum P(A_i)$
(vi) If $A \subset B$, $P(A) \le P(B)$.

3. Independence of experiments

An experiment A_r corresponds to a decomposition of the observation space E into disjoint subsets $A_{r1}, A_{r2}, \ldots A_{rK_r}$. let r = 1,2, ... n. The decompositions are <u>mutually independent</u>

provided

$$P(A_{r_1 1}\ A_{r_2 2} \ldots A_{r_n n}) = P(A_{r_1 1})\,P(A_{r_2 2}) \ldots P(A_{r_n n})$$

for any $r_1, r_2, \ldots r_n$

If $A_1, \ldots A_n$ are mutually independent, then any m of them ($m<n$) are also independent

4. Independence of events

The n events $A_1, A_2, \ldots A_n$ are mutually independent if the decompositions $E = A_k \cup \bar{A}_k$, ($k=1,2,\ldots n$) are independent. Hence the N & S conditions for the mutual independence of the events $A_1, A_2, \ldots A_n$ are the following $2^n - n - 1$ relations

$$P(A_{r_1}\ A_{r_2} \ldots A_{r_m}) = P(A_{r_1}) \ldots P(A_{r_m}),\ m = 1,2,\ldots n,$$

$$1 \leq r_1 < r_2 < \ldots < r_m \leq n.$$

Note: the independence of events in pairs does not necessarily imply their mutual independence, ie we can have $P(AB) = P(A).P(B), P(BC) = P(B).P(C)$, $P(AC) = P(A).P(C)$, but $P(ABC) \neq P(A).\ P(B)\ P(C)$.

In particular, A and B are independent if and olny if $P(AB)=P(A).\ P(B)$

5. Conditional probability

DEF. $P(B|A) = P(AB)/P(A)$

where $P(AB) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B)$

Conditional probabilities behave like probabilities :ie

$P(B|A) \geq 0$, $P(E|A) = 1$, $P(B \cup C|A) = P(B|A) + P(C|A)$ provided $BC = 0$

$$P(A|A) = 1$$

If and only if A and B are independent, $P(A|B) = P(A)$, $P(B|A) = P(B)$.

6. Random variables, or variates : A random variable is a single valued real function X (u) defined for all $u \in E$ (where E is the observation space) and for which $\{u: X(u) \leq a\} \subset \mathcal{F}$ for all real a.

We often write X for X(u).

(3) DISTRIBUTION FUNCTIONS

1. Univariate case. Take E to be the real axis, $\mathcal{F}$ the aggregate of all countable unions and intersection of subsets of E; then the non-negative completely additive set function P(A) may be defined by its values for the special intervals $(-\infty, x)$

$$P(-\infty, x) = P(X \leq x) = F(x) \quad = \text{the (cumulative) distrib. fn.},$$

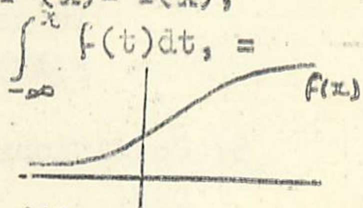
where $F(-\infty) = 0$, $F(+\infty) = 1$, F(x) is a bounded monot. non-decr. fn. since $P(a < X \leq b) = F(b) - F(a) \geq 0$ if $b > a$

At discontinuities we define $F(x) = F(x+0)$

Clearly $P(X=x) = F(x) - F(x-0)$

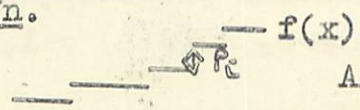
1a. Continuous type: F(x) differentiable, $F'(x) = f(x)$, the prob. (density) fn. $f(x) \geq 0$, $F(x) = \int_{-\infty}^{x} f(t)\,dt = \Pr(X \leq x)$

$\Pr(X \in dx) = \Pr(x < X \leq x+dx) = f(x)dx$



1b. Discrete type: F(x) a step-function, with jumps of magnitudes p_i at x_i $(i=1,2,\ldots)$, $\sum p_i = 1$

$\Pr(X=x_i) = p_i$, the prob. fn.



$$F(x) = \sum_{i(x)} p_i$$

where $i(x) = \{i : x_i \leq x\}$

Alternative notation $p_i = f(x_i)$

2. Bivariate case.

where $F(x,y) \geq 0$, $F(-\infty,-\infty) = F(-\infty,y) = F(x,-\infty) = F(-\infty,+\infty) = F(\infty,-\infty) = 0$

$F(\infty,\infty) = 1$

$F_{22} - F_{12} - F_{21} + F_{11} \geq 0$ where $F_{ij} = F(x_i, y_j)$ and $x_2 > x_1, y_2 > y_1$

$$P(x_1 < X \leq x_2,\ y_1 < Y \leq y_2) = F_{22} - F_{12} - F_{21} + F_{11}$$

F is monotone non-decreasing in each variable separately.
At discontinuities $F(x,y) = F(x+0,y) = F(x,y+0)$

2a. Continuous type: $\partial^2 F/\partial x \partial y = f(x,y)$ = Prob. denst. fun. of X, Y

$P(X \in dx, Y \in dy) = f(x,y)\,dx\,dy$, $F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,du\,d$

2b Discrete type: There is an enumerable set of points (x_r, y_s) and positive numbers p_{rs}; $\sum_{r,s} p_{rs} = 1$; s.t. $F(x,y) = \sum_{rs(xy)} p_{r}$

where $r\,s(xy) = \{r,s:\ x_r \leq x,\ y_s \leq y\}$ Then $P(X = x_r, Y = y_s) = p_{rs}$

3. Marginal distributions: Let (X,Y) have the d.f. $F(x,y)$, as in

The marginal d.f. of X is $F_1(x) = F(x,\infty) = P(X \leq x)$, $(= P(X \leq x, Y \leq \infty))$

of Y $F_2(y) = F(\infty,y) = P(Y \leq y)$

Then $F_1(x)$, $F_2(y)$ are univariate d.f.'s.

3a. Continuous type: If $F(x,y)$ is continuous, we define the marginal prob. fn. of X to be $f_1(x) = F_1'(x) = \int_{-\infty}^{+\infty} f(x,y)\,dy$

of Y $f_2(y) = F_2'(y)$

3b. Discrete type: If (X,Y) has discrete pr.fn. p_{rs}, the marginal prob. fn. of X is $p_1(r) = \sum_s p_{rs}$; of Y, $p_2(S) = \sum_r$

$$f_1(x_r) = \sum_s f(x_r, y_s)$$

4. Conditional distribution functions From (2), §5, we have

$P(X \leq x \mid Y \leq y) = P(X \leq x, Y \leq y) \;/\; P(Y \leq y)$

$= F(x,y) \;/\; F_2(y)$ by §3

We define $P(X \leq x \mid Y=y)$ to be $\lim_{h \to 0} P(X \leq x \mid y < Y \leq y+h)$

$= \frac{\partial F}{\partial y} / f_2(y)$ in the continuous case. The conditional prob. fn. of X, given Y=y, is then defined to be $f_{1|2}(x|y)$,

$f_{1|2}(x|y)\, dx = p(x < X \leq x+dx \mid Y=y)$

whence

$$f_{1|2}(x|y) = \frac{f(x,y)}{f_2(y)}, \text{ where } \begin{aligned} f(x,y) &= \frac{\partial^2 F(x,y)}{\partial x\, \partial y} \\ f_2(y) &= \frac{d}{dy} F(\infty, y) \end{aligned}$$

(4) INTEGRATION

We write $\int \phi(x)\, dF(x)$ to denote the Steiltjes integral of $\phi(x)$ with respect to $F(x)$.

In the continuous case, $\int \phi(x)\, dF(x) = \int \phi(x) f(x) dx$, the ordinary Riemann $\int$ ($f(x) = F'(x)$)

In the discrete case where $F(x)$ is a step function with jumps $f(x_i)$ at x_i, $\int \phi(x) dF(x) = \sum \phi(x_i) \cdot f(x_i)$ the R. integral, or sum, being taken over the appropriate range..

Note that if $F(x)$ is the d.f. of the variate X, then

$$\boxed{P(X \in A) = \int_A dF(x)}$$

(5) EXPECTATION

1. If X has d.f. $F(x)$, the expectation of any function of $\psi(X)$ of X is

$$\boxed{\mathcal{E}\psi(X) = \int \psi(x)\, dF(x)}$$ (integral taken over all possible values of X)

Similerly for bivariate distributions, using the natural generalization of § (4)

$$\mathcal{E}\psi(X,Y) = \int \psi(x,y)\, dF(x,y) = \iint \psi(x,y)\, f(x,y)\, dxdy = \sum\sum \psi(x_r,y_s)\, p_{rs}$$

Additivity: $\mathcal{E}(X+Y) = \int (x+y)\, dF(x,y)$

$$= \int x dF(x,y) + \int y\, dF(x,y) = \mathcal{E}X + \mathcal{E}Y$$

We then have : If $X \geqslant 0$, $\mathcal{E}X \geqslant 0$,

$$\mathcal{E}(aX+b) = a\,\mathcal{E}X + b$$

$$\mathcal{E}(X+Y) = \mathcal{E}X + \mathcal{E}Y$$

$\mathcal{E}(\ )$ is therefore linear operation

In particular

$$\mathcal{E}(\sum \lambda_i X_i) = \sum \lambda_i \mathcal{E}X_i$$

2. Moments If X has d.f. F(x) the r^{th} moment (about the origin) is $\mu'_r = \mathcal{E}X^r = \int x^r\, dF(x)$

the r^{th} central moment is

$$\mu_r = \mathcal{E}(X-\mu)^r = \int (x-\mu)^r dF(x), \text{ where } \mu \equiv \mu'_1$$

Conditional moments are moments of the appropriate conditional distribution

3. Variance The dispersion of a distribution may be measured by the "standard deviation", whose square is the variance, defined by (var X=) $v(X) = \mathcal{E}(X-\mu)^2 = \mathcal{E}X^2 - \mu^2$ where $\mu = \mathcal{E}(X)$

then we have: $V(X) \geqslant 0$: in fact $V(X) > 0$ unless $X \equiv$ const.

$$V(aX+b) = a^2\, V(X)$$

4. Covariance: $\mathcal{C}(X,Y) = \mathcal{E}(XY) - (\mathcal{E}X)(\mathcal{E}Y) = \mathcal{E}(X-\mathcal{E}X)(Y-\mathcal{E}Y) = \mathcal{C}(Y,X) = \text{cov}(X,Y)$

whence

$$\mathcal{C}(aX+b, cY+d) = ac\,\mathcal{C}(X,Y); \quad \mathcal{C}(X,a) = 0,$$

$$\mathcal{C}(X+U, Y+V) = \mathcal{C}(X,Y) + \mathcal{C}(U,Y) + \mathcal{C}(X,V) + \mathcal{C}(U,V)$$

Hence

$$\boxed{V(aX+bY) = a^2V(X) + 2ab\,\mathcal{C}(X,Y) + b^2V(Y)}$$

If $v(X) = \sigma_1^2$, $v(Y) = \sigma_2^2$ we define the <u>correlation</u> <u>coefficient</u> as

$$\rho(X,Y) = \frac{\mathcal{C}(X,Y)}{\sigma_1 \sigma_2}$$

whence

$$-1 \leq \rho(X,Y) \leq +1 \qquad \rho(aX+b,\ cX+d) = \rho(X,Y)$$

X, Y are uncorrelated if $\mathcal{C}(X,Y) = 0$

5. <u>Independent variates</u>. X, Y are stochastecally <u>independent</u> if $F(x,y) = F_1(x)\,F_2(y)$, $f(x,y) = f_1(x)\,f_2(y)$

F_1, F_2 are then necessarily the marginal df's (up to a const. multiplier) and f_1, f_2 the marginal probability functions. We then have $\mathcal{E}(XY) = \mathcal{E}X\ \mathcal{E}Y$ as a consequence of the def. So that independence implies uncorrelation(but not conversely).

6. MARKOFF'S INEQUALITY

If $X > 0$, and $\mathcal{E}X = \mu$ is finite then for any $k > 0$,

$$P\{X \geq k\mu\} \leq 1/k.$$

Proof: Let $\left.\begin{array}{ll} Y = 0 & \text{when } X < k\mu \\ Y = k\mu & \text{,, } X \geq k\mu \end{array}\right\}$ then $Y \leq X$ so $\mathcal{E}Y \leq \mu$

But $\mathcal{E}Y = 0.\,P(X < k\mu) + k\mu.\,P(X \geq k\mu)$, where the theorem.

7. TCHEBYCHEFF'S INEQUALITY

For any variate X with $\mathcal{E}X = \mu$ and $vX = \sigma^2 \neq 0$,

$$\boxed{P\{\,|X - \mu| \geq k\sigma\} \leq \frac{1}{k^2}} \qquad \text{for any } k > 0.$$

Proof: In Markoff's inequality replace X by $(X-\mu)/\sigma^2$

Example showing that the = signs are attainable: consider the discrete X for which $P(X=\mu) = 1-1/k^2$, $P(X=\mu \pm k\sigma) = 1/2k^2$

STANDARD DISTRIBUTIONS

The variate is X, (or X_1, X_2... in multivariate cases)

The complete specification of the probability functions listed below is

{ in the range quoted, this probability function has the value quoted, outside this range the probability function is zero.

Name	Probability function	Range
Binomial	$\binom{n}{x} p^x q^{n-x}$, $o < p < 1$, $q = 1-p$,	o,1,2,...,n
Poisson	$e^{-\mu} \mu^x / x!$, $\mu > o$,	o,1,2,...
Hypergeometric	$\binom{a}{x}\binom{k-a}{n-x} \Big/ \binom{k}{n} = \binom{n}{x} \dfrac{a^{(x)} (k-a)^{(n-x)}}{k^{(n)}}$ $o < a < k$, $o < n < k$,	o,1,2,...,n
Neg.binomial	$\binom{k+x-1}{k-1} p^k q^x$, $k > o$, $o < p < 1$, $q = 1-p$	o,1,2,...
Multinomial	$\dfrac{n!}{x_1! x_2! \ldots x_k!} p_1^{x_1} \ldots p_k^{x_k}$, $o < p_i < 1$ $\sum p_i = 1$ $\sum x_i = n$	o,1,2,...n for each x_i
Rectangular	1	$(-\frac{1}{2}, +\frac{1}{2})$
Triangular	$1 - \lvert x \rvert$	(-1,+1)
Exponential	e^{-x}	$x \geq o$
Double exponential	$\frac{1}{2} e^{-\lvert x \rvert}$	all real values
Beta	$\dfrac{1}{B(p,q)} x^{p-1} (1-x)^{q-1}$, $p > o$, $q > o$	(o,1)

Gamma	$\frac{1}{\Gamma(p)} x^{p-1} e^{-x}$, $p > 0$	$x \geq 0$
Cauchy	$1/\pi(1+x^2)$	all real values
Standard normal	$e^{-\frac{1}{2}x^2} / \sqrt{2\pi}$	,, ,, ,,
Bivariate normal	$\frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1-\rho^2}} e^{-\frac{1}{2}Q/(1-\rho^2)}$	,, ,, ,,

$$\sigma_1 > 0$$
$$\sigma_2 > 0$$
$$\rho^2 < 1$$

$$Q = \left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2$$

Standard multinormal $(2\pi)^{-\frac{1}{2}k} |\underline{V}|^{-\frac{1}{2}} e^{-\frac{1}{2}\underline{x}' V^{-1} \underline{x}}$, all real values. for each x_i

$\underline{V}$ pos.def.

Notes on the Standard distributions

1. Binomial x= number of successes in n indep. trials, where probability of a success, at any trial, is p = const.

   $\mathcal{E}X = n\,p$, $\mathcal{V}X = n\,p\,q$, $\mu_3 = n\,p\,q\,(q-p)$,

   $F(x) = I_x\,(n-x, x+1)$, (incomplete beta function ratio)

   If X binomial (n,p) and Y binomial (m,p), indep.; X+Y binomial (n+m,p)

2. Poisson (a) x = no. of occurrences of a given event in time t, where probability of a single occurrency during $\delta t = \lambda \delta t + O(\delta t)$, probability of >1 occurrences during $\delta t = O(\delta t)$, and no. of occurrences during non-overlapping time intervals are indep. of each other.

   Then x is Poisson, with parameter $\mu = \lambda t$.

(b) $\varphi(x)$ = limit of binomial prob. fn. when $n \to \infty$,

$p \to 0$, $np = \mu$

If X_1, X_2 are indep. Poisson variates with parameters

μ_1, μ_2, then $X_1 + X_2$ is Poisson $(\mu_1 + \mu_2)$

$\mathcal{E} X = \mu$, $\mathcal{V} X = \mu$; $F(x) = 1 - I_\mu(x+1)$, (incomplete

gamma fn ratio)

3. Hypergeometric x = no. of A's in a sample of n, taken

without replacement from a set of k item of which $\underline{a}$ were

A's

$$\mathcal{E} X = \frac{a}{k} n \qquad \mathcal{E} X^{(r)} = a^{(r)} \; n^{(r)} / k^{(r)}.$$

4. Negative binomial (a) $k + X$ = no. of binomial trials

(of prob. p) required to achieve k successes $\mathcal{E} X = kq/p$,

$\mathcal{V} X = kq/p^2$

If we put $q' = 1/p$, $p' = q/p$, so that $q' - p' = 1$, the

pr. fn. becomes coef. of p'^x in expansion of $(q' - p')^{-k}$,

and in this terminology we have $\mathcal{E} X = k p'$, $\mathcal{V} X = kp'q'$.

(b) x may also be regarded as a Poisson variate with varying Poisson parameter. In fact if $P(x,m) = m^x e^{-m}/x!$ while $P(m) = \alpha^\lambda m^{\lambda-1} e^{-\alpha m}/\Gamma(\lambda)$ $m > 0$ (a gamma distribution) then $P(x) = \left(\frac{\alpha}{1+\alpha}\right)^\lambda \frac{\Gamma(\lambda+x)}{x!\,\Gamma(\lambda)}(1+\alpha)^{-x}$ which is of one standard neg. binomial form with $p = \alpha/(1+\alpha)$, $q = 1/(1+\alpha)$.

5. Multinomial x_i = no. of occurrences of event A_i $(i=1,2,\ldots n)$ in k indep. trials, where at each trial $p(A_i) = p_i$ = const.
$\mathcal{E} X_i = np_i$, $\mathcal{V} X_i = np_i(1-p_i)$, $\mathcal{C}(X_i, X_j) = -n\,p_i\,p_j$

If $X_1, X_2, \ldots X_k$ are indep. Poisson variates with parameters $\mu_1, \ldots \mu_k$, then the conditional joint probability function of the X_i, given $\sum X_i = x$, is multinomial:

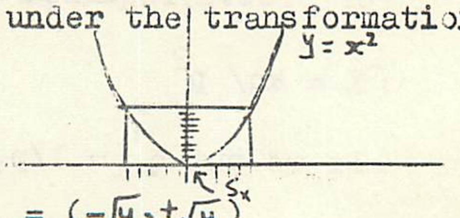
$$P(x_1,\ldots x_k \mid \sum x_i = x) = \frac{x!}{\prod x_i!}\prod\left(\frac{\mu_i}{\mu}\right)^{x_i}, \quad (\mu = \sum \mu_i)$$

Transformation of variate (Univariate case)

1. Given a variate X, d.f. $F(x)$, pr.fn. $f(x)$, and a single valued function $\theta(x)$ to find the distribution $G(y)$, $g(y)$ of

$$Y = \theta(X)$$

Let S_x be the set of all values of x which are mapped into a specified set S_y of values of y, under the transformation $y = \theta(x)$.



then $p(Y \in S_y) = p(X \in S_x)$

(eg $Y=X^2$. take $S_y = (0,y)$. Then $S_x = (-\sqrt{y}, +\sqrt{y})$

$G(y) = p(0 < Y \leq y) = p(-\sqrt{y} < X \leq +\sqrt{y})$

$= F(\sqrt{y}) - F(-\sqrt{y})$)

2. Special case where the transformation is continuous, 1-1.
$G(y) = F(\theta^{-1}(y))$ if f(x) is monotonic increasing
$= 1 - F(\theta^{-1}(y))$ " " " decreasing

In either case, if the d.f's are continuous

$|g(y)dy| = |f(x)\ dx|$ (where on r.h.s $x=x(y) = \theta^{-1}(y)$)

3. The extension to the multivariate case is straight forward. Restricting ourselves for the moment to 1-1 continuous transformation, we have for, eg, the 2 variate case,

if $U = U(X,Y)$ and $V = V(X,Y)$

then $$g(u,v) = f(x,y) \cdot \left| \frac{\partial(x,y)}{\partial(u,v)} \right|$$

Convolutions: Addition of independent variates

If X, Y are indep. variates, with d.f's $F(x)$, $G(y)$ respectively, and probability functions $f(x)$, $g(y)$ respectively, the d.f. of

$$Z = X + Y$$

is

$$H(z) = \int F(z-y)dG(y) = \int G(z-x)\ dF(x)$$

In the discrete case this becomes for the pr.fn.,

$$h(z) = \sum_y f(z-y)\ g(y) = \sum_x g(z-x)\ f(x)$$

and in the continuous case

$$h(z) = \int f(z-y)\ g(y)\ dy = \int g(z-x) f(x)\ dx$$

Exp.

1) In the following X and Y are indep. Find the distribution of $Z = X + Y$ when
   1) X is binomial (n,p), Y is binomial (m,p); [Z bin. (n+m,p)]
   2) X is Poisson (μ), Y is Poisson (λ); Z Poisson ($\lambda + \mu$)
   3) X normal (μ_1, σ_1), Y normal ($\mu_2\ \sigma_2$) ; [Z normal($\mu_1+\mu_2, \sqrt{\sigma_1^2+\sigma_2^2}$)]
   4) X rectang. ((0,1), Y the same; [Z triang.(0,2)]
   5) X gamma (p), Y gamma (q); [Z gamma (p+q)]
   6) X Cauchy (λ_1, μ_1), viz pr.fn. $\lambda/\pi \{\lambda_1^2 + (x-\mu_1)^2\}$
      Y Cauchy (λ_2, μ_2) ; [Z Cauchy ($\lambda_1 + \lambda_2, \mu_1 + \mu_2$)]

2) Deduce from 1.6 that if $X_1, X_2, \ldots X_n$ are indep., with the same cauchy distrbt., the mean $\bar{X} = \sum X_i / n$ has the same distrib. as every X_j.

3) $X_1, X_2, \ldots X_n$ are indep. standard normal variates. Prove by induction that the distribution of $Y = X_1^2 + X_2^2 + \ldots + X_k^2$ is of the gamma type, viz $e^{-\frac{1}{2}y}\ y^{\frac{1}{2}k-1} / 2^{\frac{1}{2}k}\Gamma(\frac{1}{2}k)$, $(y \geq 0)$

4) If U,V are indep. gamma variates, their sum is a gamma variate which is indep. of their quotient. Show also that ($\frac{U}{U+V}$) is a beta. (If U has pr.fn. as in Q.3, with k=p, and V ditto, with k=q, then $Y = U+V$ has ditto, with k=p+q, while $Z = U/V$ has pr.fn. Prop. to $z^{\frac{1}{2}p-1}/(1+z)^{\frac{1}{2}(p+q)}$

5) If X has the pr.fn. $\alpha^p e^{-\alpha x}\ x^{p-1}/\Gamma(p)$, $x \geq 0$, ($\alpha > 0$, $p \geq 1$), its moments are given by $\mu r = \alpha^{-r}\Gamma(r+p)/\Gamma(p)$, whence in particular $\mathcal{E}X = p/\alpha$, $\mathcal{V}X = p/\alpha^2$. (If $\alpha = 1$, $\mathcal{E}X = \mathcal{V}X$)

6) If X is normal (μ, σ), the moments of odd order r are $\mu r = \sigma^r (r-1)(r-3)\ldots 1$, (those of even order vanish)

7) Sampling from a finite population consisting of k different members:- $x_1 x_2, \dots x_k$; in which we define $\mu = \sum_1^K x_i/k$, and $\sigma^2 = \sum_1^K (x_i - \mu)^2/k$. Take a sample of size n, say $x_{r_1}, x_{r_2}, \dots x_{r_n}$, and let $m = \sum_j x_{r_j} /n$. Show that:-

$P(X_{r_i} = x_j) = 1/k$, $i = 1,2,\dots n$; $j = 1,2,\dots k$.

$P(X_{r_i} = x_p, X_{r_j} = x_q) = 1/k\,(k-1)$

$P(\text{any particular sample of } n) = 1/k^{(n)}$

$\mathcal{E} X_{r_j} = \mu$, $\mathcal{V} X_{r_j} = \sigma^2$, $\mathcal{E}\mu = \mu$,

$\mathcal{E} X^2_{r_j} = \sum_1^K x_i^2/k$, $\mathcal{C}(X_{r_i}, X_{r_j}) = -\sigma^2/(k-1)$, $\mathcal{V} M = \frac{k-n}{k-1}\frac{\sigma^2}{\mu}$

8) The continuous variate X has d.f. F(x). A new variate Y is defined by the transformation $y = F(x)$. Show that Y has the rectang. distrib$^{\text{n}}$ (o,1). ("Probability Integral Transformation")

Matrix notation

EXPECTATION

DEF: If $\underline{W}$ is a (pxq) matrix (W_{rs}), $\mathcal{E}\underline{W}$ denotes a matrix of the same size, with (r,s) element equal to $\mathcal{E}W_{rs}$.

In particular if $\underline{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix}$, $\mathcal{E}\underline{x} = \begin{pmatrix} \mathcal{E}x_1 \\ \vdots \\ \mathcal{E}x_k \end{pmatrix}$

If $\underline{A}$ is a constant matrix and $\underline{b}$ a constant vector, and $\mathcal{E}\underline{x} = \underline{\mu}$,

$$\boxed{\mathcal{E}(\underline{A}\underline{x} + \underline{b}) = \underline{A}\,\underline{\mu} + \underline{b}}$$

VARIANCE MATRIX If $\underline{x}$ is a (kx1) vector variate, its variance matrix (or variance-covariance matrix) is

$\underline{V} = \mathcal{V}\underline{x} = \mathcal{E}(\underline{x} - \underline{\mu})(\underline{x} - \underline{\mu})'$, or, equivalently,

$$\boxed{\mathcal{V}\underline{x} = \mathcal{E}\underline{x}\,\underline{x}' - \underline{\mu}\,\underline{\mu}'} \quad \text{where } \mathcal{E}\underline{x} = \underline{\mu}$$

Primes denote transposition

Hence $V_{rr} = \text{var } x_r$, $V_{rs} = \text{cov } (x_r, x_s)$.

Hence $\mathcal{V}(\underline{A}\,\underline{x} + \underline{b}) = \mathcal{V}(\underline{A}\,\underline{x}) = \underline{A}\,\underline{V}\,\underline{A}'$

Note : By def. $\mathcal{V}(\underline{x})$ is symmetric, since $\text{cov}(x_r, x_s) = \text{cov}(x_s, x_r)$.

$\mathcal{V}(\underline{x})$ is also positive -definite, provided the x_i are linearly independent in the algebraic sense, for given any vector $\underline{\lambda}$, we have $\text{var}(\underline{\lambda}'\underline{x}) = \text{var} \sum \lambda_i x_i > 0$ since variance is essentially positive, But

$$0 < \text{var}(\underline{\lambda}'\underline{x}) = \mathcal{V}(\underline{\lambda}'\underline{x}) = \underline{\lambda}'\underline{V}\,\underline{\lambda}.$$

Since the quadratic form $\underline{\lambda}'\underline{V}\underline{\lambda} > 0$ whatever the values of $\underline{\lambda}$ the result follows.

We shall frequently use the result that, if $\underline{V}$ is any pos-def. symmetric matrix, we can find(not uniquely) a nonsingular matrix $\underline{S}$ Such that $\underline{V} = \underline{S}\underline{S}'$; in particular, we can find a lower triangular matrix $\underline{S}$ with positive diagonal elements with this property, $\underline{S}$ is then unique, (Choleski),

$$\text{if } \mathcal{V}(\underline{x}) = \underline{V} = \underline{S}\,\underline{S}', \quad \mathcal{V}(\underline{S}^{-1}\underline{x}) = \underline{1}^{\delta} \quad \text{(the unit matrix)}$$

Notation : if $\underline{\lambda}$ denotes the vector $(\lambda_1, \lambda_2, \ldots \lambda_k)'$,

$\underline{\lambda}^{\delta}$ „ „ diagonal matrix $\begin{pmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \lambda_k \end{pmatrix}$

If the elements x_i of $\underline{x}$ are uncorrelated, the variance matrix of $\underline{x}$ is diagonal.

If further these x_i all have the same variances σ^2, the variance matrix is $\mathcal{V}(\underline{x}) = \underline{1}^{\delta}\sigma^2$.

<u>COVARIANCE MATRIX</u> $\quad \mathcal{C}(\underline{x},\underline{y}) = \mathcal{E}\underline{x}\,\underline{y}' - \mathcal{E}\underline{x}\,\mathcal{E}\underline{y}' = \{\mathcal{C}(\underline{y},\underline{x})\}'$

the (r,s) element is cov (x_r, y_s).

$$\mathcal{V}(\underline{x} + \underline{y}) = \mathcal{V}(\underline{x}) + \mathcal{C}(\underline{x},\underline{y}) + \mathcal{C}(\underline{y},\underline{x}) + \mathcal{V}(\underline{y}).$$

The Multivariate normal distribution

1) Let $x_1, x_2, \ldots x_k$ be k indep. standard normal variates.

their joint pr. fn. is $\prod_1^k \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} x_i^2} = (2\pi)^{-\frac{1}{2}k} \exp -\frac{1}{2} \sum x_i^2$

Introduce k new variates $y_1, y_2, \ldots y_k$ by the non-singular (and therefore 1-1) transformation

$$\underline{y} = \underline{S}\,\underline{x} + \underline{\mu}$$

then $\mathcal{E}\underline{y} = \underline{\mu}$, $\mathcal{V}\underline{y} = \underline{S}\,\underline{S}' = \underline{V}$, say.

the jacobian of the transformation is $|\partial(\underline{y})/\partial(\underline{x})| = |\underline{S}| = |\underline{V}|^{\frac{1}{2}}$

whence $|\partial(\underline{x}) / \partial(\underline{y})| = |\underline{V}|^{-\frac{1}{2}}$

Noting that $\sum x_i^2 = \underline{x}'\underline{x} = (\underline{y}-\underline{\mu})'\underline{S}^{-1'}\underline{S}^{-1}(\underline{y}-\underline{\mu}) = (y-\mu)'V^{-1}(y-\mu$

we see that the joint pr. fn. of the y_i' s is

$$f(\underline{y}) = (2\pi)^{-\frac{1}{2}k} \left|\underline{V}\right|^{-\frac{1}{2}} \exp -\tfrac{1}{2} (\underline{y}-\underline{\mu})' \underline{V}^{-1} (\underline{y}-\underline{\mu}),$$

$-\infty < \underline{y} < +\infty$

where $\underline{\mu} = \mathcal{E}\,\underline{y}$, $\underline{V} = \mathcal{V}\,\underline{y}$.

[There is no real loss of generality, and often a considerable saving of space, in taking $\underline{\mu} = \underline{0}$]

2) In the multinormal distribution, uncorrelated variates are indep.

For if $\underline{V}$ is diag, say $\underline{\lambda}^{\delta}$, f $(\underline{y})$ degenerates to

$$f(\underline{y}) = \prod_1^k \frac{1}{\sqrt{2\pi\lambda_i}} \exp -\tfrac{1}{2}(y_i - \mu_i)^2/\lambda_i$$

whence the y_i are indep. normal, with $\mathcal{E}y_i = \mu_i$, $\mathcal{V}y_i = \lambda_i$.

3) Bivariate normal distribution.

Let $\mathcal{E}y_1 = \mathcal{E}y_2 = 0$, $\mathcal{V}y_1 = \sigma_1^2$, $\mathcal{V}y_2 = \sigma_2^2$, $\mathcal{C}(y_1, y_2) = \rho\sigma_1\sigma_2$

$$\underline{V} = \begin{pmatrix} \sigma_1 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, \quad \left|\underline{V}\right| = \sigma_1^2\sigma_2^2 \; (1-\rho^2),$$

$$\underline{V}^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1/\sigma_1^2 & -\rho/\sigma_1\sigma_2 \\ -\rho/\sigma_1\sigma_2 & 1/\sigma_2^2 \end{pmatrix}$$

; hence the pr.fn. of $(y_1 y_2)$ is

$$f(y_1, y_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{(1-\rho^2)}} \exp - \frac{1}{2(1-\rho^2)} \left(\frac{y_1^2}{\sigma_1^2} - \frac{2\rho y_1 y_2}{\sigma_1\sigma_2} + \frac{y_2^2}{\sigma_2^2} \right)$$

4) If $\underline{y}$ is k-variate normal $(\underline{\mu}, \underline{V})$, ie with f$(\underline{y})$ as in § 1, and $\underline{z} = \underline{A}\,\underline{y}$ where $\underline{A}$ is nonsing, $\underline{z}$ is also k-variate normal, with $\mathcal{E}\underline{z} = \underline{A}\underline{\mu}$, $\mathcal{V}\underline{z} = \underline{A}\,\underline{V}\,\underline{A}'$.

In particular, if the y_i are indep. normals, with variance σ^2, and $\underline{z} = \underline{T}\,\underline{y}$, where $\underline{T}$ is orthogonal, then the z_i are also indep. normals, with variance σ^2.

5. Marginal distribution.

Let $\underline{y}$ be k-variate normal $(\underline{0}, \underline{V})$, $\underline{y} = \begin{pmatrix} \underline{y}_1 \\ \hline \underline{y}_2 \end{pmatrix} \begin{matrix} \updownarrow k_1 \\ \updownarrow k_2 \end{matrix}$

we seek to find the marginal pr.fn. of $\underline{y}_1$.

Let $\underline{V} = \begin{pmatrix} \underline{V}_1 & \underline{V}_2 \\ \underline{V}_2' & \underline{V}_4 \end{pmatrix}$, $\underline{V}^{-1} = \begin{pmatrix} \underline{W}_1 & \underline{W}_2 \\ \underline{W}_2' & \underline{W}_4 \end{pmatrix}$ where $\underline{V}_1$ and $\underline{W}_1$ are $(k_1 x k_1)$

whence

$V_1W_1 + V_2W_2' = 1^\delta$ so that $V_2 = -V_1W_2W_4^{-1}$

$V_1W_2 + V_2W_4 = 0$ $V_1W_1 - V_1W_2W_4^{-1}W_2' = 1^\delta$

$V_2'W_1 + V_4W_2' = 0$ $\underline{V_1^{-1} = W_1 - W_2W_4^{-1}W_2'}$

$V_2'W_2 + V_4W_4 = 1^\delta$

The marginal pr. fn. of $\underline{y}_1$ is $\int_{-\infty}^{+\infty} \cdot (k_2) \cdot \int f(\underline{y}_1, \underline{y}_2)\, dy_2$,

Now the exponent in $f(y_1, y_2)$ is proportional to

$\underline{y}' \ \underline{V}^{-1} \ \underline{y}$

| where $f(y_1, y_2) = f(\underline{y})$ is the k-variate multi norl. pr. fn. as in §1, $\mu = \underline{0}$.

$$= (y_1' \ y_2') \begin{pmatrix} W_1 & W_2 \\ W_2' & W_4 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = y_1' W_1 y_1 + \boxed{2 y_2' W_2 y_1 + y_2' W_4 y_2}$$

$$= y_1' W_1 y_1 + (y_2 + W_4^{-1} W_2' y_1)' W_4 (y_2 + W_4^{-1} W_2' y_1) - y_1' W_2 W_4^{-1} W_2' y_1$$

("Completing the square" in $\boxed{}$)

$$= y_1' (W_1 - W_2 W_4^{-1} W_2') y_1 + (y_2 + \alpha)' W_4 (y_2 + \alpha)$$, where α does not depend on y_2

$$= y_1' V_1^{-1} y_1 + (y_2 + \alpha)' W_4 (y_2 + \alpha)$$

(using result of 1st half of this paragraph)

The integration can now be carried out : the resulting marginal pr. fn. of $\underline{y}_1$ is

$$f_1 (\underline{y}_1) = A_1 \exp - \tfrac{1}{2} \underline{y}_1' \underline{V}_1^{-1} \underline{y}_1$$

where A_1 is the appropriate normal factor

which is itself k_1 - variate normal

6) Conditional distribution

Using the notation of § 5, we seek the conditional pr. fn. of $\underline{y}_2$, given $\underline{y}_1 = \underline{a}$. This is prop. to $f (\underline{a} , \underline{y}_2)$, the quad. form in the exponent of this fn. being, as in § 5,

$$\underline{y}' \underline{V}^{-1} \underline{y} = \underline{a}' V_1^{-1} \underline{a} + (y_2 + W_4^{-1} W_2' a)' W_4 (y_2 + W_4^{-1} W_2' a)$$

Whence clearly y_2 is k_2-variate multinormal with

$$\mathcal{E} \underline{y}_2 = - W_4^{-1} W_2' \underline{a} = \underline{V}_2' \underline{V}_1^{-1} \underline{a} \quad \text{(by 1st half of § 5)}$$

as the conditional expectation (whereas the unconditional expectation of $\underline{y}$ is $\underline{o}$)

and $\mathcal{V} \underline{y}_2 = W_4^{-1} = V_4 - V_2' V_1^{-1} V_2$

as the conditional variance matrix.

Generating function

Def: The g.f. of a given real sequence $(a_0, a_1, a_2, \ldots)$ is

$$A(s) = \sum_0^\infty a_r s^r$$

provided this convergies in some interval $|s| \leq s_0$.

Generating function of integral-valued variate $X \geq 0$

Let the pr. fn. of X be $Pr(X=j) = p_j$, $j = 0, 1, 2, \ldots$
and the (tail probabilities) $Pr(X > j) = q_j$, $j = 0, 1, 2, \ldots$

(Nearly $q_j = p_{j+1} + p_{j+2} + \ldots$)
with g.f's, $P(s) = \sum p_r s^r$, $Q(s) = \sum q_r s^r$. $|s| \leq 1$

Theorem

$$Q(s) = \frac{1 - P(s)}{1-s} \text{ for } |s| < 1$$

Ex Show that $\mathcal{E}X = P'(1) = Q(1)$, (provided $P'(1)$, $Q(1)$ exist)
and $\mathcal{V}X = P''(1) + P'(1) - P'^2(1) = 2Q'(1) + Q(1) - Q^2(1)$

Convolution If X, Y are independent non neg. integral valued variates, with $P_r(X=j) = a_j$, $P_r(Y=k) = b_k$,
then $P_r(X+Y=r) = a_0 b_r + a_1 b_{r-1} + a_2 b_{r-2} + \ldots + a_r b_0 = c_r$, say
the sequence $\{c_r\}$ is said to be the convolution of the sequences $\{a_r\}$, $\{b_r\}$; symbolically,

$$\{c_r\} = \{a_r\} * \{b_r\}$$

Theorem If A(s), B(s) are the g.f.'s of X, Y respectively, the g.f. of X+Y is $C(s) = A(s) . B(s)$

We extend the notation for convolutions as follows:
$\{a_r\}^{2*} = \{a_r\} * \{a_r\}$, $\{a_r\}^{3*} = \{a_r\}^{2*} * \{a_r\}$ = sequence whose g.f. is $A^3(s)$; and complete these with $\{a_r\}^{1*} = \{a_r\}$,
$\{a_r\}^{0*} = (1, 0, 0, \ldots)$ = sequence whose g.f. is $A^0(s)$.

Ex

Consider the prob. that in a sequence of Bernoulli trials the first success occurs at the $(r+1)$-th trial $(r=0,1,2,\ldots)$. This is the <u>geometric</u> distribution. Show that its g.f. is $p/(1-qs)$. Hence show that the prob. that this k-th success occurs at the $(r+k)$-th trial $(r=0,1,2,\ldots, k = \text{fixed number})$ has g.f $\left\{p/(1-qs)\right\}^k$. (This is the Pascal distribution).

Compound distribution

Let $X_1, X_2, X_3 \ldots$ be a sequence of indep. variate with a common distribution given by $P_r(X_k=j) = f_j$.

Let $S_N = X_1 + X_2 + \ldots + X_N$ Where N is a variate indep. of the X_j, with distribut. given by $P_r(N=n) = g_n$

Then
$$P_r(S_N=j) = \sum_n Pr(N=n).\, P_r(X_1 + \ldots + X_n = j)$$
$$= \sum_n g_n . \left\{f_j\right\}^{n*}, \text{ a compound distribution.}$$

Ex of compound distribution.

If the X_k have binomial distribution with $P_r(X_k=0) = q$, $P_r(X_k=1)=p$, while N has Poisson distribution with parameter λ, then $P_r(S_N=j) = e^{-\lambda p}(\lambda p)^j/j!$ i.e Poisson, with parameter λp.

(eg. S_N = no of unhealed chromosome breakages in an irradiated cell, for no. N of breakages has Poisson distribution, prob. of a breakage remaining unhealed is p, prob. of j of the N breakage remaining unhealed is binomial (N,p).)

Generating function of a compound distribution

Let $f(s) = \sum f_j x^j$ both g.f. of the X_k } and $S_N = X_1 + \ldots X_N$

$g(s) = \sum g_r s^r$ both g.f. of N)

For fixed n the g.f. of $\{f_j\}^{n*}$ is $f^n(s)$

The g.f of S_N is

$$h(s) = \sum_j Pr(S_N = j).\, s^j = \sum_j \sum_n g_n \{f_j\}^{n*} s^j$$
$$= \sum_n g_n \{\sum_j \{f_j\}^{n*} s^j\} = \sum_n g_n f^n(s)$$
$$= g(f(s)).$$

Expectation in compound distribution

$$\mathcal{E} S_N = \frac{d}{ds} \{ g(f(s)) \}_{s=1} = g'(f(1)) \cdot f'(1) = g'(1) f'(1)$$
$$= \mathcal{E} N.\, \mathcal{E} X_j$$

(Ex find the variance of S_N in terms of the expectations and variances of N and X_j)Ans.: $\mathcal{V} S_N = \sigma_N^2 \mu_X^2 + \sigma_X^2 \mu_N$

(Ex. If g_n = prob that a family has exactly n children, and the sex ratio boys: girls = p: q, (p+ q=1), then the prob. of a family having exactly j boys is the compound binomial $\sum_n g_n \binom{n}{j} p^j q^{n-j}$, with g.f. g(q+ps), where g(s)= g.f. of $\{g_n\}$

If $g_n = (1-\gamma)\gamma^n$, (geometric distrib.) show that distrib. of no. of boys in family is also geometric (as g_n with $\gamma \rightarrow \gamma p/(1-\gamma q)$

Chain reaction If we have 'generation' of particles, each having indep., fixed, prob. p_k (k=o,1,2,..) of creating k new particles; having started from a single particles in the Oth generation.

Let X_n = no. of particles in n^{th} generation, $(X_o=1)$

then $X_{n+1} = U_1+U_2 + \dots + U_{X_n}$ where each U_j has distribution $\{P_k\}$ and the U_j are mutually independent hence the g.f. of X_{n+1} is

$$P_{n+1}(s) = P_n(P_1(s)), \qquad \text{where } P_1(s) = \sum_r p_r s^r.$$

Probability x_n of termination at or before n^{th} generation:-

$$x_n = P_r(X_n=0) = P_n(0)$$

$$x_{n+1} = P_r(X_{n+1}=0) = P_{n+1}(0) = P_1(P_n(0)) = P_1(x_n)$$

thus $x_{n+1} = P(x_n)$; writing $P(s)$ for $P_1(s)$

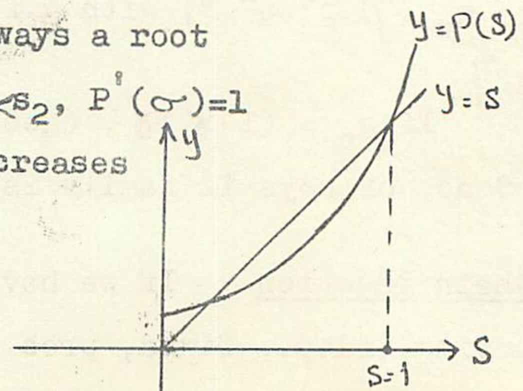
Then the prob. that this process ever terminates is $\zeta = \lim x_n$.
This certainly exist since $x_{n+1} = P(x_n) > P(x_{n-1}) = x_n$, ($P(s)$ is an incr. fn.) and so $\zeta = P(\zeta)$.

This equation may have more than one root. If so, we require the smallest root for, if η is any root other than ζ, we have $x_1 = P(0) < P(\eta) = \eta$, and by induction $x_n < \eta$, for $x_{n+1} = P(x_n) < P(\eta) = \eta$, whence $\zeta \leqslant \eta$.

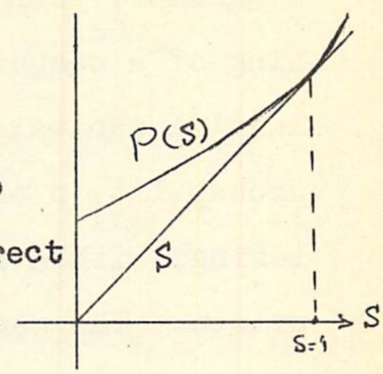
<u>Roots of $S = P(s)$</u>:- (i) $s=1$ is always a root

(ii) if s_1, s_2 are roots, $\exists\sigma$, $s_1<\sigma<s_2$, $P'(\sigma)=1$

Now this is unique, since $P'(s)$ increases steadily in $(0,1)$



Hence this can only be one pair of roots in $[0,1]$, and one of these is at $s=1$. Is there another? If so it must lie in $[0, 1]$



It is not difficult to see that there is a root $s < 1$ if and only if $P'(1) > 1$; and this root is unique. Also $P'(1) = \sum kp_k$ is the expected no. of direct descendants of a particle. Call this μ. Then we have the result:-

Let $\mu = \sum k\, p_k$ be the expected no. of direct asscendants of a single particle. If $\mu > 1$, $\exists$ a unique root $\zeta < 1$ of the equation $\zeta = P(\zeta)$, and ζ is the limit of the prob. that the process terminates after finitely many generations (Thus ζ may be interpreted as the prob. that the process will terminate, and $1-\zeta$ as the prob. of an infinitely prolonged process). If $\mu > 1$, then, $1-\zeta > 0$.

If $\mu \leq 1$, the only root of $\zeta = P(s)$) is $\zeta = 1$; so the probability tends to one that the process terminates before the n-th generation. The prob. $1-\zeta$ of infinitely prolonged process is zero.

Ex Show that expected size of n-th generation is μ^n.

The 1-dimensional random walk

1) A particle starting at $x=n$, perform a random walk with prob. p at each step of moving one unit to the right, prob. q of moving one unit to the left ($p + q=1$).

There are absorbing barriers at $x=o$, $x=a$ $(o<n<a)$.

A gambler, with initial capital £ n, plays a game(consisting of a sequence of B. trials) with an opponent when initial capital is £(a-n); at such trial our man has probability p of wining that trial, & prob. q=1-p of losing, with a stake of £1 to be won or lost as appropriate. The game continues until our man's capital is 0 (he is "runied") or is £a (he has "won").

Let prob. of his ultimate ruin, when he still has £n, be q_n

then $q_n = p\,q_{n+1} + q\,q_{n-1}$, $1 < n < a-1$

$q_1 = p\,q_2 + q$, $q_{a-1} = q.q_{a-2}$

whence

$q_n = A + B(\frac{q}{p})^n$ when $q \neq p$, $q_n = A+Bn$ when $q=p=\frac{1}{2}$

and in fact

$$\boxed{q_n = \frac{(q/p)^a - (q/p)^n}{(q/p)^a - 1}}\ ,\quad q \neq p; \qquad \boxed{q_n = 1 - \frac{n}{a}}\ ,\quad q=p=\tfrac{1}{2}$$

If p_n = prob. of this ultimate winning = prob of this oppenent's ruin, clearly p_n is obtained from q_n by $q \rightarrow p$, $p \rightarrow q$ $n \rightarrow a-n$, whence

$$p_n = \frac{(p/q)^a - (p/q)^{a-n}}{(p/q)^a - 1}\ ,\quad q \neq p; \qquad p_n = \frac{n}{a}\ ,\quad q=p=\tfrac{1}{2}$$

$$= \frac{(q/p)^n - 1}{(q/p)^a - 1} = 1 - q_n. \qquad\qquad = 1 - q_n.$$

Our gambler's expected gain is $(a-n)p_n - n\,q_n = a(1-q_n) - n$ ($= o$ if $p=q=\frac{1}{2}$) In a " fair game" ie with $p=q=\frac{1}{2}$, if he starts with £n, his prob. of winning £t before being ruined is $n/(n+t)$. (for put $a=n+t$ in above)

Effect of change of stake :

If the stake is changed from £1 to £½, this is equiv. to replacing n by 2n, and a by 2a; the prob. of ruin becomes

$$q_n^x = q_n \cdot \frac{\theta^a + \theta^n}{\theta^a + 1}, \quad (\theta = \frac{q}{p}), \text{ and this is } > q_n \text{ if } \theta > 1$$

ie if $q > p$

Thus smaller stakes increase prob. of ruin for the player having $q > p$ Thus to maximise chance of ultimately winning, stakes should be as large as possible, consistent with the target value.

<u>Expected duration of play</u>. When gambler's capital is £n, let this expected duration of play be D_n

$$D_n = p\,D_{n+1} + q\,D_{n-1} + 1, \quad 0 < n < a$$

$$D_o = o, \; D_a = D$$

The solution is

$$D_n = \frac{n}{q-p} + A + B\left(\frac{q}{p}\right)^n = \frac{n}{q-p} - \frac{a}{q-p}\,\frac{1-(q/p)^n}{1-(q/p)^a}, \quad (q \neq p)$$

$$= \quad -n^2 + A + Bn = n(a-n), \qquad (q=p=\tfrac{1}{2})$$

<u>Distribution of duration of play</u> (= duration of random walk)

Let prob. of this sum at the k^{th} step, when he starts with £n, be $u_{n,k}$

Then

$$u_{n,k+1} = p\, u_{n+1,k} + q\, u_{n-1,k}, \quad (1 < n < a-1,\ k \geqslant 1)$$

with $u_{o,k} = u_{a,k} = o, \quad k \geqslant 1$

$$u_{o,o} = 1; \quad u_{n,o} = o, \quad n > o$$

Solve by generating function $U_n(s) = \sum_n u_{n,k} s^k$

We find $\quad U_n(s) = ps\, U_{n+1}(s) + qs\, U_{n-1}(s)$

with $\quad U_o(s) = 1, \qquad U_a(s) = o$

whence

$$U_n(s) = A(s)\, \lambda_1^n(s) + B(s)\, \lambda_2^n(s), \quad \lambda_{1,2}(s) = \frac{1 \pm \sqrt{1-4pqs^2}}{2ps}$$

with

$$A(s) + B(s) = 1, \quad A\lambda_1^a + B\,\lambda_2^a = o$$

whence

$$U_n(s) = \left(\frac{q}{p}\right)^n \frac{\lambda_1^{a-n}(s) - \lambda_2^{a-n}(s)}{\lambda_1^a(s) - \lambda_2^a(s)}$$

The corresp. g. f. for the prob. of his winning at the k^{th} step is obtained from this by $p \to q$, $q \to p$, $n \to a-n$, and the g.f. for the prob. distrib. of the duration of play is the sum of these two g.f.'s

<u>Random walk with only one barrier</u> (at $x = 0$)

= game against infinitely rich opponent, $a \to \infty$

The preceding paragraph still holds, with $U_o(s) = 1$ as the role boundary condition. We find that one of the roots $\lambda(s)$ is > 1 for $o < s < 1$ and a bounded solution is obtained only if $A(s) = o$. The g. f. corresp. to $U_n(s)$ is then $V_n(s) = \lambda_2^n(s)$

This is also the g. f. of first passage times through $x = o$ of a free particle starting from $x = n$.

Markoff Chains

A Markoff chain is a sequence of states E_j in which $Pr\left\{ n^{th} \text{ state is } Ej_n/(n-1) \text{ th state was } Ej_{n-1}, (n-2) \text{ th state was } Ej_{n-2}, \ldots \text{ and } (n-r) \text{ th state was } Ej_{n-r}\right\} = Pr\left\{Ej_n/Ej_{n-1}\right\}$, for $r = 1, 2, \ldots n$, and $n = 1, 2, \ldots$

This conditional prob. that the system is in state Ej_n, given that the preceding states was Ej_{n-1} is called the transition probability from Ej_{n-1} to Ej_n and written as

$$Pr\left\{Ej_n / Ej_{n-1}\right\} = Pr\left\{Ej_{n-1} \rightarrow Ej_n\right\} = pj_{n-1}\, j_n$$

$$\text{or } Pr\left\{E_j / E_i\right\} = Pr\left\{E_i \rightarrow E_j\right\} = p_{ij} \quad (\geqslant 0)$$

Clearly

$$\sum_j p_{ij} = 1, \quad \text{all } i.$$

If we put Pr (initial state is Ej) $= a_j$ $(\geqslant o, \sum a_j = 1)$ then the probability of obtaining the sequence $Ej_o, Ej_1, Ej_2, \ldots Ej_n$ is $a_{jo}\, p_{joJ1}\, p_{j1J2} \ldots p_{jn-1Jn}$.

The system is thus completely characterized by the initial distribution $\{a_j\}$ and the matrix of transition probability $\underline{P} = (p_{ij})$. This square matrix $\underline{P}$ is subject to the restrain that $p_{ij} \geqslant o$, all i, j, and $\underline{P}'\underline{1} = \underline{1}$ (all row-sums are unity). Such a matrix is called a stochastic matrix.

<u>Higher transition probabilities:</u>

Let $p_{jk}^{(n)}$ = Pr (system is in state E_x at time r+n (was in state E_j at time r). Then, eg, $p_{jk}^{(2)} = \sum_r p_{jr}\, p_{rk}$ = (j,k) element of $\underline{P}^2$

Thus the matrix of the $p_{jk}^{(2)}$ is $\underline{P}^2$ = matrix of 2- step transition probs. Surely the matrix of n-step probabilities is $\underline{P}^n$.

<u>Absolute probabilities</u>

Pr (E_k at time n / Ej at time o) = $p_{jk}^{(n)}$, Pr (Ej at time o)= a_j,

Pr (E_k at time n) = $\sum_j a_j\, p_{jk}^{(n)}$ = k - element of the row-vector $\underline{a}'\,\underline{P}^n$

We can therefore find the stable limiting distribution (as $n \to \infty$), if any, by investigating the limiting form of $\underline{P}^n$. Now if $\underline{P}$ has latent roots λ_r, and corresponding latent vectors $\underline{x}_r$ (linearly independent) we have $\underline{P}\,\underline{x}_r = \lambda_r\,\underline{x}_r$, $\underline{P}\,\underline{X} = \underline{X}\,\underline{\Lambda}$ where $\underline{X} = (\underline{x}_1, \underline{x}_2, \ldots \underline{x}_k)$ and $\underline{\Lambda} = \text{diag}\,(\lambda_1\,\lambda_2, \ldots \lambda_k)$

Then $\underline{P} = \underline{X}\,\underline{\Lambda}\,\underline{X}^{-1}$ and $P^n = \underline{X}\;\underline{\Lambda}^n\,\underline{X}^{-1}$ where $\underline{\Lambda}^n$ = diag $(\lambda_1^{\,n}, \lambda_2^{\,n}, \ldots \lambda_k^{\,n})$

<u>Example</u> Limiting distribution in a 2-state Markoff chain

Let $\underline{P} = \begin{pmatrix} 1-\beta & \beta \\ \alpha & 1-\alpha \end{pmatrix}$, $0 < \alpha < 1$, $0 < \beta < 1$, $\alpha \neq 1 - \beta$

To find the latent roots we solve $\underline{P}\,\underline{x} = \lambda\,\underline{x}$, ie

$$\left.\begin{aligned}(1-\beta)\,x_1 + \beta\,x_2 &= \lambda x_1 \\ \alpha\,x_1 + (1-\alpha)x_2 &= \lambda x_2\end{aligned}\right\} \text{ or } \left\{\begin{aligned}(1-B-\lambda)x_1 + B\,x_2 &= 0 \\ \alpha\,x_1 + (1-\alpha-\lambda)\,x_2 &= 0\end{aligned}\right.$$

The solution must of course be arbitrary up to a constant multiplier, and will exist if $(1-p-\lambda)(1-\alpha-\lambda)-\alpha\beta = 0$

whence $\lambda^2-(2-\alpha-\beta)\lambda+(1-\alpha-\beta)=0$ [NOTE: A stochastic matrix always has a latent root equal to 1]

with roots $\lambda_1 = 1$, $\lambda_2 = 1-\alpha-\beta$

The corresp. latent vectors are $\begin{pmatrix}1\\1\end{pmatrix}$ and $\begin{pmatrix}-\beta\\ \alpha\end{pmatrix}$.

Thus

$$\underline{P}\,\underline{X} = \underline{X}\,\underline{\Lambda} \text{ becomes } \underline{P}\begin{pmatrix}1 & -\beta\\ 1 & \alpha\end{pmatrix} = \begin{pmatrix}1 & -\beta\\ 1 & \alpha\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 1-\alpha-\beta\end{pmatrix}$$

so that
$$\underline{P}^n = \begin{pmatrix}1 & -\beta\\ 1 & \alpha\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & (1-\alpha-\beta)^n\end{pmatrix}\begin{pmatrix}1 & -\beta\\ 1 & \alpha\end{pmatrix}^{-1}$$

$$= \frac{1}{\alpha+\beta}\begin{pmatrix}1 & -\beta\\ 1 & \alpha\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & (1-\alpha-\beta)^n\end{pmatrix}\begin{pmatrix}\alpha & \beta\\ -1 & 1\end{pmatrix}$$

$$= \frac{1}{\alpha+\beta}\begin{pmatrix}\alpha+\beta(1-\alpha-\beta)^n, & \beta-\beta(1-\alpha-\beta)^n\\ \alpha-\alpha(1-\alpha-\beta)^n, & \beta+\alpha(1-\alpha-\beta)^n\end{pmatrix}$$

$$= \frac{1}{\alpha+\beta}\begin{pmatrix}\alpha & \beta\\ \alpha & \beta\end{pmatrix} + \frac{(1-\alpha-\beta)^n}{\alpha+B}\begin{pmatrix}\beta & -\beta\\ -\alpha & \alpha\end{pmatrix}$$

As $n \to \infty$, $\underline{P}^n \to \frac{1}{\alpha+B}\begin{pmatrix}\alpha & B\\ \alpha & B\end{pmatrix}$, since $(1-\alpha-B)^n \to 0$

[Ex. Examin the 2-state chain with matrix $\underline{P} = \begin{pmatrix}p & q\\ p & q\end{pmatrix}$.]

Examples of transition matrices :

(1) Independent Bernoulli trials E_1 = success , E_2 = failure

$$\underline{P} = \begin{pmatrix}p & q\\ p & q\end{pmatrix}$$

(2) Random walk with absorbing barriers at $x = o$, $x = a$.

E_k mean $x = k$, $k = o, 1, 2, \dots a$.

$$p = \begin{pmatrix} 1 & o & o & o & \cdots & o \\ q & o & p & o & \cdots & o \\ o & q & o & p & \cdots & o \\ & & & & & \\ o & \cdots & o & q & o & p \\ o & \cdots & o & o & o & 1 \end{pmatrix}$$

Note the first and last rows. These character* the absorbing states represented by barriers.

Time-dependent Stochastic Processes

(1) The Birth Process   A system may be in states $E_o, E_1, E_2, \dots$ If at time t it is in state E_n, then the probability that during $(t, t+h)$ a transition occurs t E_{nn} is $\lambda h + o(h)$; the prob. of a transition t E_j $(j \neq n\ )$ is $o(h)$.

Let $P_n(t)$ = prob. that system is in state E_n at tran.t. then

$$P_n(t+h) = P_n(t).(1-\lambda_n h) + P_{n-1}(t).\lambda_{n-1} h + o(h),\ n \geq 1,$$

$$P_o(t+h) = P_o(t).(1-\lambda_o h) + o(h)$$

We now let $h \to o$, whence

$$P_n'(t) = -\lambda_n P_n(t) + \lambda_{n-1} P_{n-1}(t),\ n = o, 1, 2, \dots;\ (P_{-1} \equiv o)$$

If at time $t = o$ the system was in state E_k, we have the boundary conditions : $P_k(o) = 1$, $P_j(o) = o$ when $j \neq k$. These together with the differential-differences equation, uniquely specifying the $P_n(t)$.

Ex   The Poisson process. When $\lambda_n = \lambda$ (constant "birth rate") it is readily verified that the solution is $P_n(t) = (\lambda t)^n e^{-\lambda t}/n!$ (taking initial state to be E_0). Here we say that system-∞ in state E_n at time t when n is the number of occurrences of no n chosen event during time t; the average rate of recurrence being λ. In a population of reproducing entities, n is the no. of birth; there is a const. prob. λh of <u>one</u> birth during a short time h.

<u>Ex</u> The Yale process. If there is a prob. $h + o(h)$ that any member of the population gives birth (to a single of during time h, then when population size is n the prob. of a single additional member appearing m (t, t +h) is $n\lambda h + o(h)$. Thus $\lambda_n = n\lambda$, and $P_n'(t) = -n\lambda P_n(t) + (n-1)\lambda P_{n-1}(t)$ if the initial population size is k, then

$$P_n(t) = \binom{n-1}{n-k} e^{-k\lambda t}(1-e^{-\lambda t})^{n-k}, \quad n \geq k$$

(2) <u>Birth and Death Process</u> We consider a system as in (1) above, but with the additional possibility of a transition $E_n \Rightarrow E_{n-1}$, $(n \geq 1)$, with prob. $p_n h + o(h)$ of such a transition during (t, t+ h), and prob o(h) of a transition $E_n \rightarrow E_{n \pm j}$, $j \neq 1$.

We have

$$P_n(t+h) = P_n(t)\left\{1 - \lambda_n h - p_n h\right\} + \lambda_{n-1} h\, P_{n-1}(t) + \mu_{n+1} h\, P_{n+1}(t) + o(h)$$

where

$$P_n'(t) = -(\lambda_n + \mu_n) P_n(t) + \lambda_{n-1} P_{n-1}(t) + \mu_{n+1} P_{n+1}(t), \quad n \geq 1$$

with $P_0'(t) = -\lambda_0 P_0(t) + \mu_1 P_1(t)$

<u>Ex</u> Simplified telephone trunking problem: Assume an infinite

number of channels available, & prob. of a used time being fixed during $(t, t+h)$ is $\mu h + o((h)$. (Then leads to " exponental holding time", viz prob. that holding time exceeds $t =$ prob. that time is not fixed during $(o,t) = e^{-\mu t}$).

Assume also that the prob. of an incoming call commencing during $(t, t+h)$ is $\lambda h + o(h)$ and that the prob. of more than one call is $O(h)$. Then the number of incoming calls commencing during (o,t) has a Poisson distribution with expectation λt. ("Poisson input with parameter λ").

We say the system is in state E_n if n lines are ingaged Then the prob. of one line being freed during $(t, t+h)$ is $n\mu h + o(h)$ while prob. of one new line being required is $\lambda h + o(h)$.

Thus we have a birth and death process with

$$\lambda_n = \lambda , \mu_n = n\mu,$$

whence

$$P_n'(t) = -(\lambda + n\mu) P_n(t) + \lambda P_{n-1}(t) + (n+1)\mu P_{n+1}(t), \ n \geq 1 \quad (A)$$
$$P_o'(t) = -\lambda P_o(t) + \mu P_1(t)$$

We may nearly find the expected value $m(t)$ of n at time t, for $m(t) = \sum n P_n(t)$, so $m'(t) = \lambda - \mu m(t)$, from the diff.equs whence

$$m(t) = A e^{-\mu t} + \frac{\lambda}{\mu} = (r - \frac{\lambda}{\mu}) e^{-\mu t} + \frac{\lambda}{\mu} \quad \text{(if inital value is r)}$$
$$\to \lambda/\mu \quad \text{as } t \to \infty$$

Similarly we may find the second moment, and hence the variance (Alternatively we might obtain a (partial) differential equation for the generating function and solve that.)

Extension: If the number of channels available is finite, and equal to a, say; - then as long as there is at least one channel free, we are in the situation discussed above. Beyond this, however, we may have people waiting for a line.

Let E_n denotes the state where n = total no. of people either being served or waiting: when $n > a$ there are $n-a$ people waiting. When $n \leqslant a$ the governing equations are(A) above, but when $n > a$

$$P_n'(t) = -(\lambda + a\mu) P_n(t) + \lambda P_{n-1}(t) + a\mu \ P_{n+1}(t) \quad \text{(B)}$$

We shall consider only the limit: $\lim_{t \to \infty} P_n(t) = \phi_n$, say, obtainable by putting $P_n'(t) = 0$ in (A) and (B) (thus is the state of "statistical equilibrium").

From (A) we get $\lambda p_o = \mu p_1$ and $(\lambda + n\mu) p_n = \lambda p_{n-1} + (n+1)\mu p_{n+1}$

whence

$$p_n = \frac{(\lambda/\mu)^n}{n!} p_o \quad \text{for} \quad n \leqslant a \ ;$$

from (B) we get $(\lambda + a\mu) p_n = \lambda p_{n-1} + a\mu \ p_{n+1}$

whence

$$p_n = \frac{(\lambda/\mu)^n}{a!\, a^{n-a}} \cdot po \qquad \underline{\text{Note}} \ \sum \frac{p_n}{p_o} \ \text{converges}$$

if $\frac{\lambda}{\mu} < a$

ORDER STATISTICS

I. Suppose X has the c.d.f. $F(x) = Pr\{X < x\}$, and that the derivative $F'(x)$ exists everywhere. Suppose also that we have n independent observations on X which when arranged in order of magnitude, are

$$x_1 < x_2 < \ldots\ldots < x_i < \cdots < x_j < \ldots\ldots < x_n$$

To obtain the p.d.f. of x_i, we argue as follows:

Choose any one of the observations, before they are made (e.g. the first observation to be taken, or the second, etc..). The probability that this is in the small interval (x_i, x_i+dx_i) is $F'(x_i)dx_i$. This chosen observation will in fact be the i th in order of magnitude if, out of the remaining (n-1) observations, (i-1) are less than x_i, the rest being greater than x_i. By Bernouilli the probability of this is

$$\frac{(n-1)!}{(i-1)!\,(n-i)!}\left\{F(x_i)\right\}^{i-1}\left\{1-F(x_i)\right\}^{n-i}$$

Since the original choice can be made in n ways, the probability that any one of the observations is x_i in (x_i, x_i+dx_i), and that this observation is the ith in order of magnitude is

$$\phi(x_i)\,dx_i = n\frac{(n-1)!}{(i-1)!(n-i)!}F'(x_i)\left\{F(x_i)\right\}^{i-1}\left\{1-F(x_i)\right\}^{n-1}dx_i \qquad (1)$$

This gives the required p.d.f.

It is clear that

$$\int_{-\infty}^{\infty}\phi(x_i)dx_i = \frac{n!}{(i-1)!(n-i)!}\int_{-\infty}^{\infty}F'(x_i)\left\{F(x_i)\right\}^{i-1}\left\{1-F(x_i)\right\}^{n-i}dx_i$$

$$= \frac{n!}{(i-1)!\,(n-i)}\int_{0}^{\infty}y^{i-1}\left\{1-y\right\}^{n-i}dy$$

on putting

$$y = F(x_i)$$

Using the fact that

$$\int_0^1 t^r(1-t)^s \, dt = r\,!\; s\,! \;/\; (r+s+1)! \qquad (2)$$

we conclude that $\int_{-\infty}^{\infty} \phi(x_i) \; dx_i \;=\; 1$

II. For the joint p.d.f of x_i and x_j $(i \leqslant j)$, we choose any two (in $n(n-1)$ ways) to lie respectively, in $(x_i, x_i + dx_i)$ and in $(x_j, x_j + dx_j)$, with probability $F'(x_i)\, F'(x_j)\, dx_i\, dx_j$.

The remaining $(n-2)$ observations must be distributed with $(i-1)$ of them less than x_i, with $(j - i - 1)$ of them between x_i and x_j, and the remaining $(n - j)$ must be greater than x_j.

Hence

$$\phi(x_i, x_j) = n(n-1) \frac{(n-2)!}{(i-1)!\,(j-i-1)!\,(n-j)!} \; F'(x_i) F'(x_j) \left\{F(x_i)\right\}^{i-1} \cdot \left\{F(x_j) - F(x_i)\right\}^{j-i-1} \left\{1 - F(x_j)\right\}^{n-j} \qquad (3)$$

and it should be noted that $\phi\,(x_i, x_j) = o$ for $x_j < x_i$

III. <u>Application on the Rectangular distribution</u>

If X has a rectangular distribution in the range $(o,1)$, we have in this range $F(x) = x$ and $F'(x) = 1$ so that

$$\phi(x_i) = \frac{n!}{(i-1)!\,(n-i)!} \quad x_i^{\,i-1} \quad (1-x_i)^{n-i}$$

and

$$\phi(x_i, x_j) = \frac{n!}{(i-1)!\,(j-i-1)!\;(n-j)!} \; x_i^{\,i-1} (x_j - x_i)^{j-i-1} (1-x_j)^{n-j}$$

$$= o \quad \text{when } x_j < x_i$$

On using (2), we find

$$\mathcal{E}(x_i) = \int_0^1 x_i \, \phi(x_i) dx_i = \frac{n!}{(i-1)!\,(n-i)!} \cdot \frac{i\,!\;(n-i)!}{(n+1)!} = \frac{i}{n+1} = \frac{i}{m}$$

where $m = n + 1$

and

$$\mathcal{E}(x_i\, x_j) = \iint\limits_{o<x_i<x_j<1} x_i\, x_j\, \phi(x_i, x_j) dx_i\, dx_j$$

$$= \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \int_o^1 x_j(1-x_j)^{n-j} \left\{ \int_o^{x_j} x_i^{\,i}(x_j - x_i)^{j-i-1} dx_i\, dx_j \right\}$$

$$= i(j + 1) \;/\; (n + 1)\;(n + 2)$$

$$= i(j + 1) \;/\; m(m + 1)$$

Hence

$$\mathcal{C}(x_i, x_j) = \mathcal{E}(x_i\, x_j) - (\mathcal{E}x_i)(\mathcal{E}x_j) = \frac{i(j+1)}{m(m+1)} - \frac{ij}{2}$$

$$= \frac{i(m - j)}{m^2(M + 1)}$$

IV Standard error of quantiles.

Among the various quantities measuring location and dispersion, there is one group, namely the quantiles, which are not algebraic functions of the observations and whose sampling variances cannot accordingly be determined by the ordinary methods. Order statistics results are very useful here.

Equation (1) expresses the distribution of x_i, the member of the sample below which a proportion $\frac{i}{n}$ of the members fall i.e. the ith quantile

Put $\quad i = nq$

So that $\quad n - i = n(1-q) = np$ say

The distribution (1) has a model value given by differentiating $\phi\,(x_i)$ with respect to x_i and equating to zero.

This gives

$$(i-1)\;\frac{F'(x_i)}{F(x_i)} - (n-i)\;\frac{F'(x_i)}{1-F(x_i)} + \frac{F''(x_i)}{F'(x_i)} = o \qquad (4)$$

This equation being satisfied by the model value x'. Now for large n, the factor $\frac{F''}{F'}$ will in general be small compared with other terms in (4), i and n-i being large. We may therefore neglect it and (4) becomes to order n^{-1},

$$\frac{q}{F} - \frac{p}{1 - F} = 0$$

or

$$F(x\ ) = q \qquad (5)$$

Now let us investigate the distribution (1) in the neighbourhood of the model value .

Put

$$F = q + \zeta$$

(1) becomes (neglecting constants)

$$(q + \zeta)^{nq} (p - \zeta)^{np}$$

Taking logarithms and expanding we have (except for constants)

$$nq \log (1 + \zeta/q) + np \log (1 - \zeta/p)$$

$$= nq \quad (\zeta/q - \tfrac{1}{2} \zeta^2/q^2 + \ldots) + np\ (-\zeta/p - \tfrac{1}{2} \zeta^2/p^2 \ \ldots)$$

$$= -n \zeta^2/2\, pq + \text{terms of higher order in } \zeta$$

Thus for large samples, the distriubiton of ζ is

$dP \propto \exp (- n \zeta^2/2\, pq)\, d\zeta$

or evaluating the necessary constant

$$dP = \frac{1}{\sqrt{2}\sqrt{(\frac{pq}{n})}} \exp \left(\frac{-n}{2pq} \zeta^2 \right) d\zeta \qquad (6)$$

(6) shows that ζ in the limit is distributed normally with variance

$$v\,(\zeta) = pq/n \qquad (7)$$

To find the variance of x_i we note that

$$d\,\zeta = d\ F\,(x_i) = f\ dx_i$$

Hence $\quad v(x_i) = pq/n\ f^2\,(x_i)$

If x_i is the median, then

$$p = q = \tfrac{1}{2}$$

and we have

$$v(\text{median}) = \frac{1}{4\,n\,f^2}$$

where f is the median ordinate

ex: If the parent population is normal the median ordinate is $\frac{1}{\sigma}$ 0.39894, σ^{2} being the variance of the parent Hence the standard error of the median is

$$\frac{\sigma}{\sqrt{n}} \cdot \frac{1}{2 \times 0.39894}$$

$$= 1.2533 \frac{\sigma}{\sqrt{n}}$$

This is considerably bigger than the standard error of the normal sample mean $(\sigma/\sqrt{n})$

V. Distribution of the range

It is obvious from (3) that the joint p.d.f. of the maximum and the minimum is given by

$$\phi(x_1,x_n)\,dx_1\,dx_n = n(n-1)F'(x_1)\,F'(x_n)\left\{F(x_n) - F(x_1)\right\}^{n-2}\,dx_1\,dx_2$$

$$= 0 \quad \text{for } x_n < x_1 \qquad (8)$$

If the original distribution is rectangular with a range (o,1) then

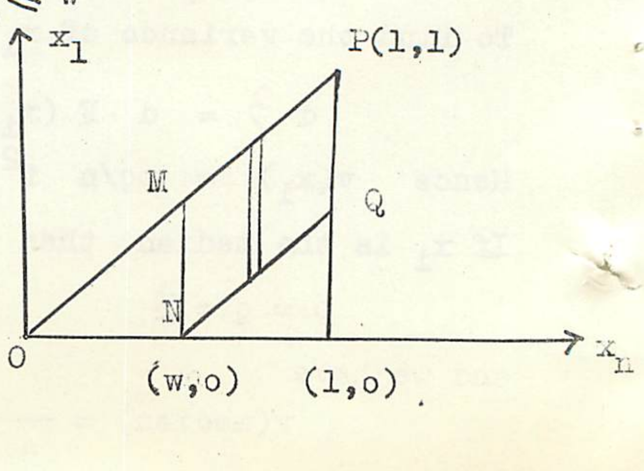
$$\phi(x_1,x_n)\,dx_1\,dx_n = n(n-1)\,(x_n-x_1)^{n-2}\,dx_1\,dx_n \qquad (9)$$

$$= 0 \quad \text{for } x_n < x_1$$

The c.d.f, of the range w is then given by integrating (9) over the values of x_1, x_n such that $x_n - x_1 \leqslant w$

i.e. over the area bounded by the two lines $x_n = x_1$

$x_n = x_1 + w$



This area could be divided into the two areas oNM & NQPM

Hence

$$F(w) = \int_{o}^{w}\int_{x_1=o}^{x_n} \phi(x_1,x_n)\, dx_1\, dx_n + \int_{x_n=w}^{1}\int_{x_1=x_n-w}^{x_n} \phi(x_1,x_n)dx_1\, dx_n$$

$$= w^n \qquad\qquad + n\, w^{n-1}\,(1-w)$$

$$= n\, w^{n-1} - (n-1)\, w^n$$

Hence

$$f(w)\, dw = n(n-1) \left\{ w^{n-2} - w^{n-1} \right\} \qquad o \leq w \leq 1$$

It is easy to verify that

$$E\,(w) = (n-1) \,/\, (n+1)$$

$$V\,(w) = 2(n-1) \,/\, (n+1)^2\, (n+2)$$

Equation (8) is soluble explicitly in some other particular cases. The normal case has been fairly completely studied by Tippet and Pearson. Tippet found the first four moments of the distribution of the range, tabulated the mean values for values of n up to 1000 and gave a diagram for determining the standard errors. These tables and diagrams are reproduced in "Tables for Statistioians and Biometricians" Part II.

APPENDIX

Problems

Probability Theory

1. Of three independent events the chance that the first only should happen is a; the chance of the second only is b; the chance of the third only is C. show that the independent chances of the three events are respectively

$$\frac{a}{a+x}, \quad \frac{b}{b+x}, \quad \frac{c}{c+x},$$ where x is a root of the equation $x^2 = (a+x)(b+x)(c+x)$?

2. A box contains ten pairs of gloves. A draws out a single glove: then B draws one; then A draws a second then B draws a second. Show that A's Chance of drawing a pair is the same as B's and that the chance of neither drawing a pair is 290/323.

3. Two bags A and B each contain N balls; the balls in A are white and in B are Black. A ball is drawn simultaneously from each bag; the ball taken from A is put into B, and vice versa. This procedure is carried out n times in all. What is the expected number of white balls in A after the n^{th} trial?

4. Under the conditions of a certain game to be played by X and Y, player making the first move in a particular game has a 5/8 the chance of wining that game. No games can be drawn. The player who wins a game has the right to make the first move in the next game.

X is to make the first move in the first of a series of n games. What is the number of games that X may expect to win during this seris?

5. A and B shoot at mark, and A hits it once in n times, and B once in (n-1) times. If they shoot alternatively, A commencing, compare their chances of first hitting the mark.

6. In a game of squash racquets between two players, the server scores one point and serves again if he wins a rally, but if he loses the rally the score remains the same and the other player serves.

Assuming that the two players A and B, are equally matched and that there is no advantage in service, find the probability that r rallies will be played without the score changing and hence the expectation of r.

If A is serving, what are the probabilities that (i) A will be the first to gain a point and (ii) B will be the first to gain a point?

7. A game is played in which 3 dice are thrown simultaneously If the numbers on the upper faces sum to less than 12, the thrower puts a number of pence into the central bank corresponding to the difference between 12 and the number he has thrown. If he throws exactly 12 he pays nothing; if he throws over 12 he takes from the bank pence equal to difference between the number he has thrown and 12. Two men, one with 3 dice and one the banker play the game. What is the expectation of the bankers gain?

8. There are three urns labelled 1, 2, and 3. The first contains 5 white balls and one black ball, the second contains 4 white and 2 black balls, the third contains 3 white and 3 black balls. The balls are indistinguishable except as regards color. A ball is taken from Urn 1 and mixed with the balls

of urn 2. A ball is taken from urn 2 and mixed with the balls of Urn 3. Finally a ball is taken from Urn 3 and mixed with the balls of Urn 1. This circular procedure is carried out three times in all.

What is the prob. that the number of white balls in Urn 1 is K at the end of this sets of operations.

9. Two packs of cards are made up in the following way. The first pack consists of the suits of heart, diamonds and clubs, and the second pack consists of the suits of clubs, spades and diamonds. A Sampling experiment is carried out which consists of drawing a card from the first pack; if this card is red a second card is drawn from the same pack, if it is black the second card is drawn from the second pack. The colour of the second card is noted, both cards are replaced and the procedure is repeated. If n trials are made, calculate the probability distribution of k, the second number of red cards seen at the second draw.

10. The target on a bombing range is a circle of c feet radius. Jone, a bomb aimer, may be assumed always to aim at the centre of the target and to make independent errors in line and range, each with the same standard deviation σ. In order to pass out as a qualified man it is necessary for him to hit the target at least k times in n trips, the bombs being dropped single on each trip. If $C=25$, $\sigma = 50$, $n=30$, $K=15$, Calculate Jones chances of successding in this test.

11. Two men A and B play a single game of tennis. If the probability of winning any point is P, show that his probability of winning the game is

$$P^4 (1 + 4q + 10 q^2) + \frac{20 p^5 q^3}{q^2 - p^2}$$

12. In an Urn there are nxk balls. The balls are indistingishable except as regards colour; there are k different colours and n balls of each. A ball is drawn and replaced after its colour has been noted. Show that the probability that r drawings will be required in order that all colours will have been seen at least once is

$$\left\{(k-1)^{r-1} - {}^{k-1}C_1 \ (k-2)^{r-1} + {}^{k-1}C_2 \ (k-3)^{r-1} \ \ldots\right\} \ /k^{r-1}$$

13. Part of the retina of the eye may be regarded as a lattice of sensitive points (rods) on which light falls so that one photon of hight hits one, and only one, rod.

Consider as a simplified model, a square lattice of n^2 rods. Two photons of light fall at random on this lattice fitting different rods, what is the probability that the two rods hit are adjacent (where a rod not at the edge of the lattice is regarded as "adjacent" to 8 others)?

14. An experiment is carried out on the effects of coal dust in the air upon mice. Assuming that the probability that a mouse, which is alive at the beginning of the, t^{th} month, will die during the month is p_t. find the expected length of life of a mouse.

The experiment started with 60 mice but at the end of each month two of the surviving mice were removed from the experiment for examination and their lengths of survival are unknown. It was found that 15 mice died in the first month, 18 in the second, 13 in the third, 5 in the fourth and the remaining mouse died in the fifth month. Obtain an estimate of the expected length of life.

15. In a certain process, events occur randomly in time; i.e. the probability of an event occuring in the small interval $(t, t+\delta t)$ is $\lambda \delta t$ for some $\lambda > 0$, independently of what has happened previously. Show that the time interval between two successive events is a random variable with an exponential distribution, and the number of events occuring in a period of fixed length is a poisson variable. What is the relation between the average interval and the average number of events in $(0, T)$?

16. A trafic light has a constant probability λdt of changing to green after being red or to red after being green in any infinitesmal interval dt. Show that a car arriving at a random instant has a probability ½ of passing through without waiting and a probability element $\frac{1}{2}\lambda \exp(-\lambda w)$ of waiting a time ω) where $\omega > 0$

Distributions, Moments, and Moment Generating function

1. A variable y is defined as

$$y = x + z$$

where x is a discrete variable, distributed according to the Poisson law with mean λ, and z is a continuous variable, independent of x with probability density function

$$p(z) = 1 \qquad 0 \leq z \leq 1,$$
$$= 0 \qquad \text{elsewhere}$$

Derive the first four moments of y
If $\lambda = 1$, obtain the vlaue of y_0 where

$$P\{y > y_0\} = 0.05$$

2. Screws manufactured by a certain machine are passed through two gauges, the first of which automatically rejects all screws whose length is less than ℓ_1 and the second reject those whose length is greater than ℓ_2. It is found that, out of N screws tested, n_1, have lengths less than ℓ_1 and n_2 have lengths greater than ℓ_2.

Assuming that length is approximately normally distributed, estimate the mean and standard deviation of length of all screws produced.

Derive an estimate of the mean length of those screws which are not rejected by the gauges.

Obtain the two estimates of mean length in the case where $N=100$, $n_1 = 9$, $n_2 = 3$, $\ell_1 = 96$ mm. and $\ell_2 = 104$ mm.

3. u is a unit normal variable, and χ^2 is distributed as follows

$$P(\chi^2) = \frac{1}{2^{f/2}\,\Gamma(f/2)} (\chi^2)^{f/2-1} e^{-\frac{1}{2}\chi^2} \qquad 0 < \chi^2 < \infty$$

independently of u. Write down the joint distribution of u and χ^2. A transformation is made to new variables t and s, given by $t = \frac{u\sqrt{f}}{\sqrt{\chi^2}}$ and $s = \frac{1}{2}\chi^2$. Write down the equations of the inverse transformation. Find the Jacobians

$$\frac{\partial(t,s)}{\partial(u,\chi^2)} \quad \text{and} \quad \frac{\partial u, \chi^2}{(t\,,\,s)} \quad \text{and}$$

verify that there product is unity. Hence find the joint distribution of t and s, and by integrating out s derive the distribution t.

(you will need a further transformation $s\,(1+ t^2/f = v)$.

4. A random variable x is known to have the distribution

$$p(x) = c\,(1+ \frac{x}{a})^{m-1}\; e^{-mx/a}\,, \quad -a \leq x \leq \infty$$

Find the constant c and the first four moments of x. Derive the linear relation between the β_1 and the β_2 of this distribution. Sketch p(x) for m=1,2.

5. Define: (a) a random variable, (b) a characteristic random variable (c) an elementary probability law (d) a cumulative probability law. Use the technique of the characteristic random variable (or any other method) to establish the first four moments of a Poisson distribution. x and y are two random variable which each follow a Poisson law. How would you use a pair of observed values of x and y to test the hypothsis that the expections of the two Poissons distributions are the same?

6. In sample of size n from a bivariate Normal population, the sample correlation coefficient r has (approximately) mean ρ and variance $(1-\rho^2)^2/n-1$. Derive a transformation $Z=f(r)$ which will give a statistic Z with variance approximately independent of ρ and n and give this variance.

7. The distribution of a random variable x is such that $y=\log x$ is normally distributed with mean μ and standard deviation σ. Derive an expression for the rth moment of x about zero in terms of M and σ. Hence show that the coefficient of variation of x is independent of μ. Comment on possible practical applications of this result.

8. Derive the binomial distribution of probabilities, stating clearly the assumptions involved. Obtain the first four moments of the distribution and show that under certain conditions the values of β_1 and β_2 tend to those of the normal distribution.

9. Given that, if χ^2 is the sum of squares of f independent unit normal variables, its sampling distribution is

$$P(\chi^2) = \frac{1}{2^{\frac{1}{2}f}\,\Gamma(\frac{1}{2}f)} (\chi^2)^{\frac{1}{2}f-1} e^{-\frac{1}{2}\chi^2}$$

find the first two moments of χ^2

A sample of size $n(x_1, x_2, \ldots, x_n)$ is drawn at random from a normal population with mean ξ and standard deviation. Derive the first two moments of

$$\theta = \sum_{i=1}^{n} (x_i - \bar{x})^2$$ and compare with those of χ^2.

10. For the distribution

$$dF = \frac{\Gamma\{\frac{1}{2}(n+1)\}}{\sqrt{\pi}\,(\frac{1}{2}n)} \operatorname{sech}^n x \, dx, \quad -\infty \leq x \leq \infty$$

Show that

$$\mu_{2r}(n) = \frac{n^2}{(2r+1)(2r+2)} \{\mu_{r+2}(n) - \mu_{2r+2}(n+2)\},$$

$$r \geqslant 0$$

Hence show that

$$\mu_{2r+2}(n) = (2r+1)(2r+2) \left\{ \frac{\mu_{2r}(n)}{n^2} + \frac{\mu_{2r}(n+2)}{(n+2)^2} + \frac{\mu_{2r}(n+4)}{(n+4)^2} + \ldots \right.$$

For the variance show that

$$\frac{1}{n} < \mu_2 < \frac{1}{n-2}$$

11. State and prove the binomial theorem in probability. Demonstrate the correction between the sum of a number of binomial probabilities and the incomplete Beta Function ratio viz.,

$$\int_0^x t^{r-1} (1-t)^{s-1} \, dt \Big/ \int_0^1 t^{r-1}(1-t)^{s-1} \, dt$$

$$0 \leq x \leq 1$$

Explain over what range of values this relationship is useful in practice and what procedure you would adopt outside the range of the tables.

12. given that $f_1(t)=f_2(t) = 1-|t|$ when $|t|\leq 1$, while $f_1(t)=0$ when $|t|>1$, show that it is possible to have

$$X_1 + X_2 = Y_1 + Y_2$$

13. Find the first four moments of the type I distribution

$$P(x) = \frac{1}{B(r,s)} x^{r-1} (1-x)^{s-1} \qquad 0<x<1$$

Show that (i) If t has the distribution P(t)=

$$\frac{1}{\sqrt{\nu}\, B(\tfrac{1}{2},\tfrac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{\frac{-\nu+1}{2}}$$

then

$$u = \frac{1}{1+ t^2/\nu} \text{ is a type I variable.}$$

(ii) If F has the distribution $P(F) = \dfrac{\nu_1^{\nu_1/2}\,\nu_2^{\nu_2/2}}{B(\frac{\nu_1}{2},\frac{\nu_2}{2})} F^{\nu_1/2-1}$

$$\{\nu_1 F + \nu_2\}^{-\frac{\nu_1+\nu_2}{2}}, \quad F \geq 0$$

then $\frac{1}{F}$ has a distribution of the same form as F with ν_1, ν_2 interchanged.

(iii) $w = \nu_1 F / \nu_1 F + \nu_2$ is a type I variable.

14. The moment statistic m_r is defined as $\sum (x-\bar{x})^r/n$. Show that in normal samples

$$\mathrm{Corr}\,(m_2, m_4) = \frac{\sqrt{3}\,(n-1)}{\sqrt{4n^2 - qn+6}}$$

15. A quantity of bacteria is treated with a disinfectant, but owing to the method of application not all bacteria receive treatment. It is necessary that of each bacterium receives t treatments in order to be killed. If the probability of any one bacterium being treated in any one application of the disinfectant is constant and equal to p, calculate the proportion of deaths after the nth application, where $n \geq t$. What is the generating function for those proportions?