THE ANALYSIS OF RANGE WITH OUTPUT LINEARLY DEPENDENT UPON STORAGE

By

Mirko J. Melentijevich

September 1965

HYDROLOGY PAPERS
COLORADO STATE UNIVERSITY
Fort Collins, Colorado

11

THE ANALYSIS OF RANGE WITH OUTPUT LINEARLY DEPENDENT UPON STORAGE

by

Mirko J. Melentijevich

HYDROLOGY PAPERS
COLORADO STATE UNIVERSITY
FORT COLLINS, COLORADO

September 1965 No. 11

ACKNOWLEDGMENTS

An expression of gratitude is given by the writer to those responsible for his support through a graduate research assistantship while at Colorado State University, to the National Science Foundation for the support of the project through research grants, and to the National Center of Atmospheric Research, Boulder, Colorado, for the use of its CDC 3600 digital computer system.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

LIST OF FIGURES AND TABLES (Continued)

ABSTRACT

The objectives of this investigation are twofold: (i) Determination of the distribution function for the cumulative sums S_t when the output is dependent upon these sums; and (ii) Development of probability expressions for the range of cumulative departures of a stochastic variable.

The basic relationship between input, output and cumulative sum is expressed by the equation of continuity

$$\frac{dS}{dt} = q_i(t) - \alpha S; \text{ with } \alpha = \text{constant.}$$

The following principal assumptions concern the fluctuating portion $q_i(t)$: (i) $q_i(t)$ is a normal independent variable with a mean of zero and a variance σ^2; (ii) The correlation between the values of $q_i(t)$ at different times t_1 and t_2 exists only when $|t_1 - t_2|$ is very small; (iii) $q_i(t)$ varies an extreme rapid amount when compared with the variation of the cumulative sum S_t.

Theoretical equations and hypothesis have been substantiated by the data generation method, which employs a digital computer. The large generated sample for computations consisted of 100, 000 normal independent numbers, with mean zero and variance unity.

The large amount of data agreement between the data generation method and that obtained from the theory indicates the validity of the theoretical equations.

The distribution function for the cumulative sums S_t when output is dependent upon those sums is defined by

$$f(S, t; S_o) = \frac{\alpha^{1/2}}{[\pi(1-e^{-2\alpha t})\sigma^2]^{1/2}} e^{-\alpha(S-S_o e^{-\alpha t})^2/(1-e^{-2\alpha t})\sigma^2}$$

The equations for the expected values and variance of range, R_n, are derived as

$$E[R_n] = \frac{(1+3\alpha^2 e^{-2\alpha})}{\sqrt{\alpha\pi}} \sum_{t=1}^{n} \frac{\sqrt{1-e^{-2\alpha t}}}{t}, \text{ and}$$

$$\text{Var}[R_n] = (\ln 2 - 2/\pi)(1-8\alpha\ e^{-20\alpha})\left(\frac{1-e^{-2\alpha}}{\alpha^2}\right) \sum_{t=1}^{n} \frac{1-e^{-2\alpha t}}{t}.$$

THE ANALYSIS OF RANGE WITH OUTPUT LINEARLY DEPENDENT UPON STORAGE

By: Mirko J. Melentijevich

CHAPTER I

MATHEMATICAL MODELS

1. Introduction. Assume the availability of a record $[X_k]$ of mutually independent random variables with a common distribution f(x). The mean for these variables is assumed zero. Let $S_n = X_1 + X_2 + \cdots + X_n$ and let

$$M_n = \max \quad [0, S_1, S_2, \ldots, S_n],$$
$$m_n = \min \quad [0, S_1, S_2, \ldots, S_n]. \qquad 1.1$$

The random variable M_n is the maximum surplus of the cumulative sums, S_i, with $i = 0, \ldots, n$. The random variable m_n is the maximum deficit of the cumulative sums, S_i, and the random variable

$$R_n = M_n - m_n \qquad 1.2$$

is the range. These values are shown in fig. 1.1 where 0t is the time axis and 0S the axis of the cumulative sums of mutually independent random variables with means zero. The curve representing these sums is 0ABC, and 0C is the time of n units.

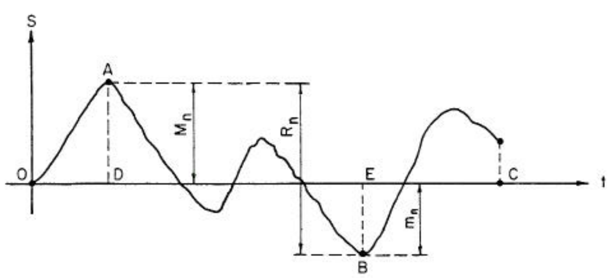

Fig. 1.1

For the cumulative sums, S_i, the maximum positive sum DA is the maximum surplus, M_n; the maximum negative sum EB is the maximum deficit, m_n; and the sum DA + EB is the range, R_n. At present very few theoretical results are available for the characteristics of surplus, deficit and range. The results available are usually valid only for $[X_k]$ normally distributed, with mean zero and variance σ^2. Feller [4], in 1951, derived results for the asymptotic distribution of the range of the cumulative departures of a stochastic variable from its mean. In terms of the water storage-water yield relation in hydrology, his results apply to cases where the variable is the annual flow of a stream, with the annual draft equal to the mean annual flow, and a lengthy time period. In particular, Feller obtained:

$$E[R_n] = 1.60 \sqrt{n}, \quad \text{and}$$
$$\mathrm{Var}[R_n] = 0.2181\, n. \qquad 1.3$$

He assumed that $[X_k]$ is a sequence of mutually independent random variables with $E[X_k] = 0$, and $\mathrm{Var}[X_k] = 1$. Feller's asymptotic solution depends on the variance of S_n alone.

A. A. Anis and E. H. Lloyd [1, 1953] solved the planning storage capacity problem of a reservoir when the water storage distribution over a given number, n, of years is not known. The storage after r years may be regarded as the sum of r annual increments. This real problem may be approximated by an ideal which has annual increments that are independent variables with a common distribution. Applications to other storage problems are obvious. Anis and Lloyd derived the expected value of the range over n years, as

$$E[R_n] = \sqrt{\frac{2}{\pi}} \sum_{r=1}^{n=1} r^{-1/2} \qquad 1.4$$

In conclusion, they noted that the asymptotic value of the range for large n is

$$2\sqrt{\frac{2}{\pi}\, n} \qquad 1.5$$

which is an agreement with Feller's results.

Thomas and Fiering [7] analyzed the following: (i) four record lengths (10, 25, 50, and 100 years) for streamflow distributions (Normal, Gamma with skewness -0.5, 0.5 and 1.0); (ii) three serial correlation coefficients in annual flows (0.0, 0.1 and 0.2); and (iii) three degrees of regulation (100 percent, 90 percent and 80 percent. Since the number of parameter combinations increases in a multiplicative fashion, 144 combinations defined the sample space for their investigation. In each instance, the annual flows are assumed to be derived from a population with the mean unity and the standard deviation 0.25. A digital computer was coded to generate a varied number of idential sets of data; for each combination they chose 100. The most important result of their study was the verification of Hurst's [5] and Feller's [4] theory with a constant outflow equaling the mean inflow. The agreement for small values of n is fairly close to Hurst's and Feller's, and becomes extremely close as n increases. Computations based on the formula given by Anis and Lloyd are only slightly better for n = 10 than those given by Hurst and Feller. The results of the study by Thomas and

Fiering indicate that the variance of the basic variable is by far the most important parameter, even for relatively short record lengths. The relative unimportance of skewness of the distribution of basic variables is clearly demonstrated. The serial correlation tends to increase the required storage. This fact is also supported by the queueing theory.

2. Distribution function for cumulative sums, S_i, with output dependent upon those sums. The basic relation between input, output and cumulative sum, S_i, is expressed by the equation of continuity which for an incompressible medium may be written as

$$\frac{dS}{dt} = Q_i(t) - Q_o(t) \qquad 1.6$$

where $Q_i(t)$ is the input, $Q_o(t)$ is the output, and dS/dt is the rate of change in time of the cumulative sum S_i. The symbol t relates here to a continuous time series. When a discrete time series is used, the symbol n replaces t.

Here the input is taken as an independent stochastic variable, while the output is a dependent variable.

They are expressed here as:

$$Q_i(t) = Q_a + q_i(t) \qquad 1.7$$

with Q_a = average input, and $q_i(t)$ = fluctuating deviations of input about Q_a;

$$Q_o(t) = Q_a + q_o(t) \qquad 1.8$$

with Q_a = average output, and $q_o(t)$ = fluctuating deviations of output about Q_a. The basic proportionality of output to the cumulative sum, S, is

$$q_o(t) = \alpha S, \qquad 1.9$$

with α = constant, $-\infty < \alpha < \infty$; and $-\infty < S < \infty$.

Equations 1.7 through 1.9 give the basic mathematical model for the case studied in this paper as:

$$\frac{dS}{dt} = q_i(t) - \alpha S. \qquad 1.10$$

Equation 1.10 is a Langevin equation for the Brownian motion of a free particle. As for the fluctuating part $q_i(t)$ the following principle assumptions are made for this equation: (i) The mean of $q_i(t)$ is zero, or $\overline{q_i(t)} = 0$; (ii) The correlation between the values of $q_i(t)$ at different times t_1 and t_2 exists only when $|t_1 - t_2|$ is very small; (iii) $q_i(t)$ varies extremely rapidly compared to the variation of S; and (iv) $q_i(t)$ is normally distributed with mean zero and variance σ^2. The problem is to determine the probability at which the cumulative sum, S, after the time t lies between S and S + dS, with $S = S_o$ at $t = 0$ being the initial sum (or storage in the case of reservoirs).

The Langevin equation has been solved by many authors, using different integration methods. One solution of this equation was given by S. Chandrasekhar [8]. His method of solution is used here to obtain a solution of eq. 1.10.

Consequently, "solving" eq. 1.10 should be understood in the sense of specifying a probability density distribution $f(S, t; S_o)$. Physical circumstances of the problem require that $f(S, t; S_o)$ follow a distribution which is independent of S_o as $t \to \infty$:

$$f(S, t; S_o) = \frac{1}{(2\pi \sigma_s^2)^{1/2}} e^{-S^2/2\sigma_s^2} \qquad 1.11$$

with σ_s^2 the variance of S for $t \to \infty$. This requirement on $f(S, t; S_o)$ conversely requires that $q_i(t)$ satisfy certain statistical conditions. The general solution of eq. 1.10 is:

$$S - S_o e^{-\alpha t} = e^{-\alpha t} \int_o^t e^{\alpha \xi} q_i(\xi)\, d\xi \qquad 1.12$$

Consequently, the statistical properties of

$$S - S_o e^{-\alpha t} \qquad 1.13$$

must be the same as those of

$$e^{-\alpha t} \int_o^t e^{\alpha \xi} q_i(\xi)\, d\xi. \qquad 1.14$$

As $t \to \infty$, eq. 1.13 tends to S; hence the distribution of

$$\lim_{t \to \infty} \left\{ e^{-\alpha t} \int_o^t e^{\alpha \xi} q_i(\xi)\, d\xi \right\} \qquad 1.15$$

must be the distribution

$$\frac{1}{(2\pi \sigma_s^2)^{1/2}} e^{-S^2/2\sigma_s^2} . \qquad 1.16$$

The right-hand side of eq. 1.12 may be written for $\xi = j\Delta t$ as

$$e^{-\alpha t} \sum_j e^{\alpha j \Delta t} \int_{j\Delta t}^{(j+1)\Delta t} q_i(\xi)\, d\xi. \qquad 1.17$$

Let

$$q(\Delta t) = \int_t^{t+\Delta t} q_i(\xi)\, d\xi, \qquad 1.18$$

and the physical meaning of $q(\Delta t)$ is that it represents the input during an interval Δt. Equation 1.12 then becomes

$$S - S_o e^{-\alpha t} = \sum_j e^{\alpha (j\Delta t - t)} q(\Delta t) \qquad 1.19$$

with the condition that the quantity on the right-hand side tends to the distribution eq. 1.16 as $t \longrightarrow \infty$. This further requires that the probability of occurrence of different values of $q(\Delta t)$ be governed by the distribution function

$$f\left[q(\Delta t)\right] = \frac{1}{(2\pi \sigma^2 \Delta t)^{1/2}} e^{-\left|q(\Delta t)\right|^2/2\Delta t \sigma^2} \qquad 1.20$$

where

$$\sigma^2 = 2\alpha \sigma_S^2 \quad \text{or} \quad \sigma_S^2 = \frac{\sigma^2}{2\alpha} . \qquad 1.21$$

To prove this assertion the distribution function $f(S, t; S_o)$ derived on the basis of eqs. 1.19 and 1.20, does in fact tend to the distribution eq. 1.16 as $t \longrightarrow \infty$. Let

$$S = \int_o^t \theta_1(\xi)\, q_i(\xi)\, d\xi . \qquad 1.22$$

Then, the probability distribution of S is given by

$$f(S) = \frac{1}{\left[2\pi\sigma^2 \int_o^t \theta_1^2(\xi)\, d\xi\right]^{1/2}} e^{-S^2/2\sigma^2 \int_o^t \theta_1^2(\xi)\, d\xi} . \qquad 1.23$$

In order to prove this, the interval (o, t) is first divided into a large number of subintervals of duration Δt, so that

$$S = \sum_j \theta_1(j\Delta t) \int_{j\Delta t}^{(j+1)\Delta t} q_i(\xi)\, d\xi . \qquad 1.24$$

Using eq. 1.19, S can be expressed in the form

$$S = \sum_j s_j, \qquad 1.25$$

where

$$s_j = \theta_1(j\Delta t)\, q(\Delta t). \qquad 1.26$$

According to eq. 1.20, the probability distribution of s_j is given as:

$$f(s_j) = \frac{1}{[2\pi\, \theta_1^2(j\Delta t)\, \sigma^2 \Delta t]^{1/2}} e^{-s_j^2/2\theta_1^2(j\Delta t)\sigma^2} \qquad 1.27$$

Hence,

$$f(S) = \frac{1}{[2\pi \sigma^2 \sum_j \theta_1^2(j\Delta t)\,\Delta t]^{1/2}} e^{-S^2/2\sigma^2 \sum_j \theta_1^2(j\Delta t)} \qquad 1.28$$

As

$$\sum_j \theta_1^2(j\Delta t)\,\Delta t = \int_o^t \theta_1^2(\xi)\, d\xi , \qquad 1.29$$

then,

$$f(S) = \frac{1}{[2\pi\, \sigma^2 \int_o^t \theta_1^2(\xi)\, d\xi\,]} e^{-S^2/2\sigma^2 \int_o^t \theta_1^2(\xi)\, d\xi} \qquad 1.30$$

which proves eqs. 1.22 and 1.23.

The right-hand side of eq. 1.12 may be expressed as

$$\int_o^t \theta_1(\xi)\, q_i(\xi)\, d\xi \qquad 1.31$$

with

$$\theta_1(\xi) = e^{\alpha(\xi - t)} . \qquad 1.32$$

With the foregoing definition of $\theta_1(\xi)$, eq. 1.30 governs the probability distribution of

$$S - S_o e^{-\alpha t} . \qquad 1.33$$

Since

$$\int_o^t \theta_1^2(\xi)\, d\xi = \int_o^t e^{2\alpha(\xi - t)} d\xi = \frac{1}{2\alpha}\left(1 - e^{-2\alpha t}\right), \qquad 1.34$$

and taking into account the relationship shown in eq. 1.21 then

$$f(S, t; S_o) = \frac{1}{[2\pi(1 - e^{-2\alpha t})\sigma_S^2]^{1/2}} e^{-(S - S_o e^{-\alpha t})^2/2(1 - e^{-2\alpha t})\sigma_S^2} \qquad 1.35$$

Therefore, eq. 1.35 converges to

$$f(S, t; S_o) = \frac{1}{(2\pi\sigma_S^2)^{1/2}} e^{-S^2/2\sigma_S^2} \qquad 1.36$$

for $t \longrightarrow \infty$.

This proves the assertion made that with the statistical properties of $q(\Delta t)$ implied in eqs. 1.20 and 1.21, eq. 1.19 leads to a distribution $f(S, t; S_o)$ which tends to be independent of S_o as $t \longrightarrow \infty$.

3. Fokker-Planck partial differential equation for cumulative sums, S. The second method for deriving eq. 1.35 is by adopting the Fokker-Planck partial differential equation for cumulative sums. For this equation $f(S, t; S_o)$ is the fundamental solution.

When t increases by Δt, S will increase by ΔS in the distribution function $f(S, t; S_o)$.

Let the probability for an increase between the limits ΔS and $\Delta S + d(\Delta S)$ be $f(\Delta S, S, t)\, d(\Delta S)$, then

$$f(S+\Delta S, t+\Delta t; S_o) = \int_{-\infty}^{+\infty} f(S, t; S_o)\, f(\Delta S, S, t)\, d(\Delta S). \qquad 1.37$$

Suppose that the probability of an increase, ΔS, is independent of the fact that for $t = o$, $S = S_o$, then the integrand for powers of ΔS is

$$\begin{aligned} f(S, t; S_o)\, f(\Delta S, S, t) = {} & f(S+\Delta S, t)\ f(\Delta S, S+\Delta S, t) - \\ & - \Delta S\,[f'(S, t; S_o)\, f(\Delta S, S, t) + \\ & + f(S, t; S_o)\, f'(\Delta S, S, t)] + \\ & + \frac{(\Delta S)^2}{2}[f''(S, t; S_o)\, f(\Delta S, S, t) + \\ & + 2f'(S, t; S_o)\, f'(\Delta S, S, t) + \\ & + f(S, t; S_o)\, f''(\Delta S, S, t)] + \dots \end{aligned} \qquad 1\ 38$$

The resulting integrals all have simple meanings. For instance

$$\int_{-\infty}^{+\infty} f(\Delta S, S+\Delta S, t)\ d(\Delta S) = 1, \qquad 1.39$$

$$\int_{-\infty}^{+\infty} \Delta S\, f(\Delta S, S, t)\, d(\Delta S) = [\Delta S], \qquad 1.40$$

$$\int_{-\infty}^{\infty} \Delta S^2\ f''(\Delta S, S, t)\ d(\Delta S) = \frac{\partial^2[\Delta S^2]}{\partial(S+\Delta S)^2}, \qquad 1.41$$

and so on. Developing the left hand side in powers of Δt by using

$$\left.\begin{aligned} \lim_{\Delta t \to o} \frac{[\Delta S]}{\Delta t} &= f_1(S+\Delta S, t); \\ \lim_{\Delta t \to o} \frac{[\Delta S^2]}{\Delta t} &= f_2(S+\Delta S, t); \end{aligned}\right\} \qquad 1.42$$

and assuming that

$$\lim_{\Delta t \to o} \frac{[\Delta S^k]}{\Delta t} = 0; \quad \text{for } k > 2. \qquad 1.43$$

then for S replacing $S + \Delta S$

$$\begin{aligned} \frac{\partial f(S, t; S_o)}{\partial t} = {} & \\ & \frac{1}{2} f_2(S, t)\frac{\partial^2 f(S, t; S_o)}{\partial S^2} + \left[\frac{\partial f_2(S, t)}{\partial S} - f_1(S, t)\right]\frac{\partial f(S, t; S_o)}{\partial S} + \left[\frac{1}{2}\frac{\partial^2 f_2(S, t)}{\partial S^2} - \frac{\partial f_1(S, t)}{\partial S}\right] f(S, t; S_o). \end{aligned} \qquad 1.44$$

The function $f_1(S, t)$ and $f_2(S, t)$ must be determined in order to verify the assumption of eq. 1.43. It comes from the storage equation that

$$\Delta S = -\alpha\, S\, \Delta t + \int_t^{t+\Delta t} q_i(\xi)\, d\xi. \qquad 1.45$$

As the mean of $q_i(\xi)$ is zero, then

$$[\Delta S] = -\alpha\, S\, \Delta t = -\alpha\,(S+\Delta S)\,\Delta t, \qquad 1.46$$

Which is obtained by neglecting the higher power terms of Δt. From this

$$\lim_{\Delta t \to o} \frac{[\Delta S]}{\Delta t} = f_1(S+\Delta S, t) = -\alpha\,(S+\Delta S). \qquad 1.47$$

In the same way, by neglecting the correlation between the values $q_i(\xi)$ at ξ_1 and ξ_2

$$[\Delta S^2] = \sigma^2\, \Delta t. \qquad 1.48$$

so that:

$$f_2(S+\Delta S, t) = \sigma^2 = 2\,\alpha\,\sigma_s^2 = \text{constant}. \qquad 1.49$$

All the powers of ΔS greater than one become proportional to like powers of Δt, so that eq. 1.43 is satisfied. Therefore,

$$\frac{\partial f(S, t; S_o)}{\partial t} = \alpha\frac{\partial}{\partial S}\left[S\, f(S, t; S_o)\right] + \frac{\sigma^2}{2}\frac{\partial^2 f(S, t; S_o)}{\partial S^2} \qquad 1.50$$

which is the required Fokker-Planck partial differential equation, of which eq. 1.35 is then the fundamental solution.

4. <u>The range of the cumulative sums, S_i, with output dependent upon those sums.</u> Let $[q_k]$ be a sequence of mutually independent random variables with a common density function $f(q)$, with $E[q_k] = 0$ and $\mathrm{Var}[q_k] = \sigma^2$. The basic eq. 1.10 given in finite differences form

$$S_k - S_k - 1 = q_k - \alpha\ \frac{S_k + S_{k-1}}{2} \qquad 1.51$$

$$S_k = \frac{2}{2+\alpha}\ q_k + \frac{2-\alpha}{2+\alpha}\ S_{k-1} \qquad 1.52$$

Equation 1.52 gives further for $S_o = 0$;

$$S_o = 0;$$

$$S_1 = \frac{2}{2+\alpha} q_1;$$

$$S_2 = \frac{2}{2+\alpha} q_2 + \frac{2(2-\alpha)}{(2+\alpha)^2} q_1;$$

$$S_3 = \frac{2}{2+\alpha} q_3 + \frac{2(2-\alpha)}{(2+\alpha)^2} q_2 + \frac{2(2-\alpha)^2}{(2+\alpha)^3} q_1;$$

$$\cdots\cdots\cdots \qquad 1.53$$

$$S_n = \frac{2}{2+\alpha} q_n + \frac{2(2-\alpha)}{(2+\alpha)^2} q_{n-1} + \frac{2(2-\alpha)^2}{(2+\alpha)^3} q_{n-2} + \frac{2(2-\alpha)^3}{(2+\alpha)^4} q_{n-3} + \cdots$$

where α is a constant.

The maximum surplus of the cumulative sums, S_i, for $i = 0, 1, \ldots, n$ is then, according to eq. 1.1, $M_n = \max[0, S_1, S_2, \ldots, S_n]$ and the maximum deficit is $m_n = \min[0, S_1, S_2, \ldots, S_n]$. The range is $R_n = M_n - m_n$.

The sums, S_i, are asymptotically normally distributed and, therefore, the asymptotic distribution of the range is independent of the function $f(q)$. The sum S_n can then be considered as the value at time $t = n$ of a continuously changing normal variable S_t. According to eq. 1.35 with $S_o = 0$, S_t is a normal variable with mean zero and variance $\frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha t})$. As $t \longrightarrow \infty$, S_t approaches a normal variable with mean zero and variance $\sigma^2/2\alpha$. It should be noted that the term $(1 - e^{-2\alpha t})$ is larger than 0.99 for $t > 2.303/\alpha$.

(a) <u>Mean and variance of range.</u> The mean range is $E[M_n - m_n] = E[M_n] - E[m_n]$. Let $y_n(x)$ be the probability density function of M_n and let $\phi_n(x)$ and $1 - \psi(x)$ be the distribution function, respectively, of M_n and m_n, so that

$$\phi_n(x) = P_r(M_n \leq x) \text{ and } \psi(x) = P_r(m_n \geq x). \qquad 1.54$$

Let $G_n(S_1, S_2, \ldots, S_n)$ be the joint distribution function of $(S_1, S_2, \ldots, S_n)$, so that

$$\phi_n(y) = \int_{-\infty}^{y} \cdots \int_{-\infty}^{y} G_n(S_1, S_2, \ldots, S_n)\, dS_1\, dS_2 \ldots dS_n \qquad 1.55$$

and

$$\psi(-y) = \int_{-\infty}^{y} \cdots \int_{-\infty}^{y} G_n(S_1, S_2, \ldots, S_n)\, dS_1\, dS_2 \ldots dS_n \qquad 1.56$$

corresponding to the observation that $E[M_n] = -E[m_n]$ for symmetrical input distributions. The function $\phi_n(y)$ and $\psi(-y)$ are thus integrals of the same type. If $f(x)$ is an even function it follows that $\phi_n(y) = \psi(-y)$.

For $\alpha \neq o$ it is very difficult to find the exact analytical expression for the joint distribution function $G_n(S_1, S_2, \ldots, S_n)$. It is also practically impossible to derive general theoretical expressions for the moments of range. If one accepts the hypothesis that the mean and the variance of range depend on the variance of S_t alone, for a given α, the following expressions are tested by the writer on a digital computer by the data generation method (Monte Carlo method):

$$E[R_n] = C_1 \sum_{t=1}^{n} \frac{\sqrt{\mathrm{Var}[S_t]}}{t} \qquad 1.57$$

and

$$\mathrm{Var}[R_n] = C_2 \sum_{t=1}^{n} \frac{\mathrm{Var}[S_t]}{t}. \qquad 1.58$$

The constants C_1 and C_2 are functions of α alone. Using the computer, for nine different values of α $(-0.04 \leq \alpha \leq 2.00)$, the expected values and variance of range were calculated for n between two and fifty. From these results it is found that the variations of C_1 and C_2 may be approximated by the following expressions:

$$C_1 = \sqrt{\frac{2}{\pi}}\,(1 + 3\alpha^2 e^{-2\alpha}) \qquad 1.59$$

and

$$C_2 = 4(\ln 2 - 2/\pi)(1 - 8e^{-20\alpha})\frac{(1 - e^{-2\alpha})}{2\alpha}. \qquad 1.60$$

Assuming that $[q_k]$ is a normal variable with mean zero and variance unity, the following equations for the expected values and variance of range are obtained

$$E[R_n] = \frac{(1 + 3\alpha^2 e^{-2\alpha})}{\sqrt{\alpha\pi}} \sum_{t=1}^{n} \frac{\sqrt{1 - e^{-2\alpha t}}}{t}; \qquad 1.61$$

and

$$\mathrm{Var}[R_n] = (\ln 2 - 2/\pi)(1 - 8\alpha e^{-20\alpha})\left(\frac{1 - e^{-2\alpha}}{\alpha^2}\right) \sum_{t=1}^{n} \frac{1 - e^{-2\alpha t}}{t}. \qquad 1.62$$

As α tends to zero, the expected values of range are the same as found by Anis and Lloyd [1] and the variance has the same form as derived by Feller [4], i.e.,

$$E[R_n] = \sqrt{\frac{2}{\pi}} \sum_{t=1}^{n} \frac{1}{\sqrt{t}}; \qquad 1.63$$

and

$$\mathrm{Var}[R_n] = 4(\ln 2 - 2/\pi)n. \qquad 1.64$$

As α tends to infinity the expected values of range and variance converge to zero.

(b) Correlation coefficient between M_n and m_n. The expression for the correlation coefficient between M_n and m_n is derived from the general equation for the variance of R_n,

$$\mathrm{Var}[R_n] = \mathrm{Var}[M_n] + \mathrm{Var}[m_n] - 2\,\mathrm{Cov}[M_n, m_n] \tag{1.65}$$

$$\mathrm{Cov}[M_n, m_n] = \rho\,(M_n, m_n)\sqrt{\mathrm{Var}[M_n]}\;\sqrt{\mathrm{Var}[m_n]} \tag{1.66}$$

$$\mathrm{Var}[m_n] = \mathrm{Var}[M_n], \text{ and} \tag{1.67}$$

$$\rho\,(M_n, m_n) = 1 - \frac{\mathrm{Var}[R_n]}{2\,\mathrm{Var}[M_n]}. \tag{1.68}$$

For n large, and $\alpha = 0$, $\mathrm{Var}[R_n] = 0.2181n$, eq. 1.31, and $\mathrm{Var}[M_n] = n\left(1 - \frac{2}{\pi}\right) - \frac{2+\sqrt{2}}{\pi}\sqrt{n}$, given by Anis [2]. For this case the correlation coefficient between M_n and m_n becomes:

$$\rho\,(M_n, m_n) = 1 - \frac{0.2181}{2(1 - 2/\pi)} = 0.700. \tag{1.69}$$

This value is verified by using the large amount of random numbers simulated on a digital computer. For $\alpha \neq 0$ the correlation coefficient between M_n and m_n was obtained by using only random numbers and it was shown to be less than 0.700. If $\alpha \longrightarrow \pm\infty$ the correlation coefficient approaches zero.

CHAPTER II

TESTING AND IMPROVING MATHEMATICAL MODELS USING THE DATA GENERATION METHOD

1. Boundary conditions. Equations and hypotheses derived in the previous chapter have been proven by using the Monte Carlo or the data generation method with simulation of a large amount of random numbers on a digital computer. The data used consisted of 100,000 random numbers of an independent normal variable with mean zero and variance unity. The program was such as to use random numbers in blocks of 100, with a total of 1000 groups. The cumulative sums of this variable are computed. The probability density distributions for the accumulated sum during the period of n units, the accumulated sum at the time n, the range, the upper maximum sum, and the lower minimum sum, are obtained for the time lags n = 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 45 and 50. The first four moments are determined, both about the origin and about the mean. The variance, the standard deviation, the coefficient of variation, the skew coefficient and the excess are also computed for distributions of each of the above statistics. All calculations were made for the following nine values of the basic parameter α: -0.040, -0.020, 0.000, 0.040, 0.100, 0.200, 0.400, 0.800 and 2.000.

The analysis of the problem with the output being a linear function of the cumulative sums shows different results depending on the boundary conditions taken for the equation

$$S_k = c_1 S_{k-1} + c_2 X_k, \qquad 2.1$$

with c_1 and c_2 dependent on the following boundary conditions.

Three cases for the integration of the above equation are considered:

(1) Case I. The instantaneous output depends on the instantaneous value of cumulative sum, S, with $c_1 = (2 - \alpha)/(2 + \alpha)$, and $c_2 = 2/(2 + \alpha)$;

(2) Case II. The output at the time n depends on the value of cumulative sum, S, at the beginning of an interval, with $c_1 = 1 - \alpha$, and $c_2 = 1.00$; and

(3) Case III. The output at the time n depends on the value of cumulative sum, S, at the end of an interval, with $c_1 = c_2 = 1/(1 + \alpha)$.

Only Case I produced the same results as the theoretical examples derived in the previous chapter.

Figure 2.1 shows how the constants c_1 and c_2 change with α for the three above cases. The differences between these three cases increases with an increase of α.

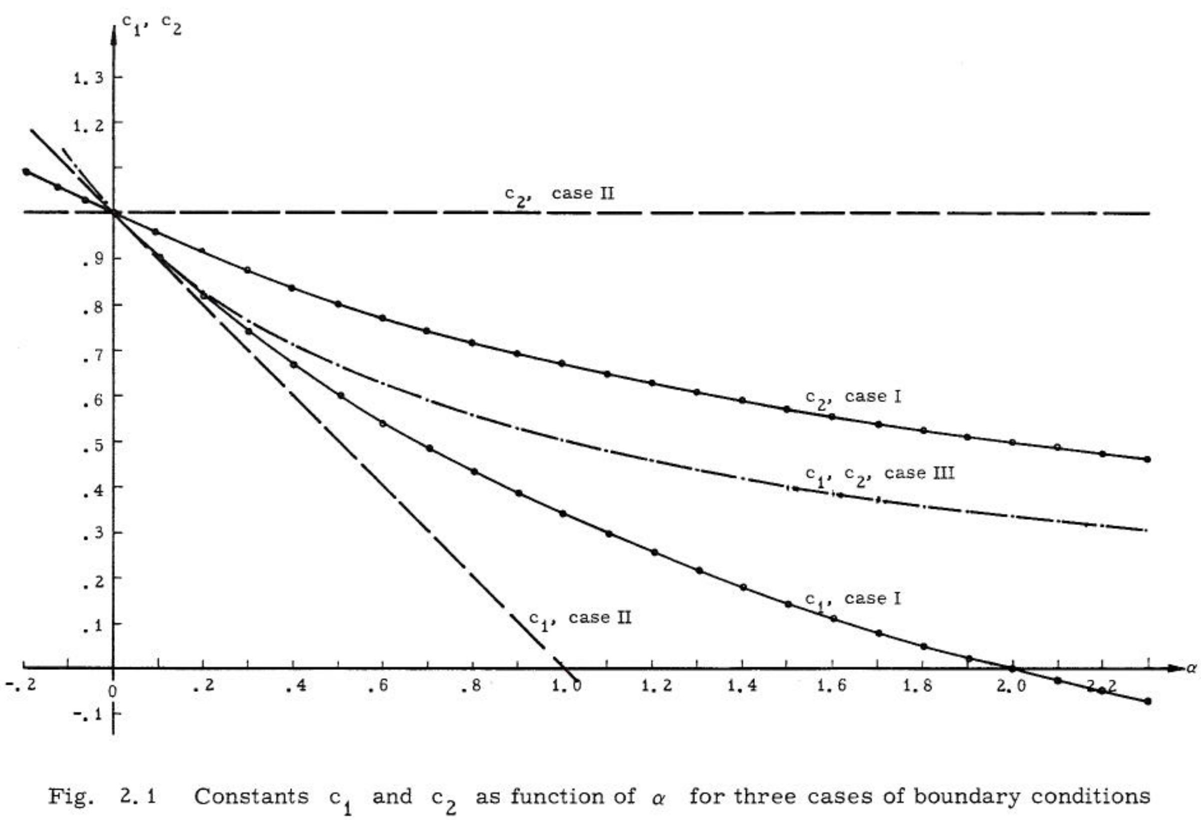

Fig. 2.1 Constants c_1 and c_2 as function of α for three cases of boundary conditions

The expression $1 - e^{-2\alpha t}$ which appears in almost all equations of the previous chapter is presented in fig. 2. 2. as a function of α, for various values of t, with $t = n\Delta t$ and $\Delta t = 1$, or a selected unit period. Therefore, t and n are interchangeable in this text.

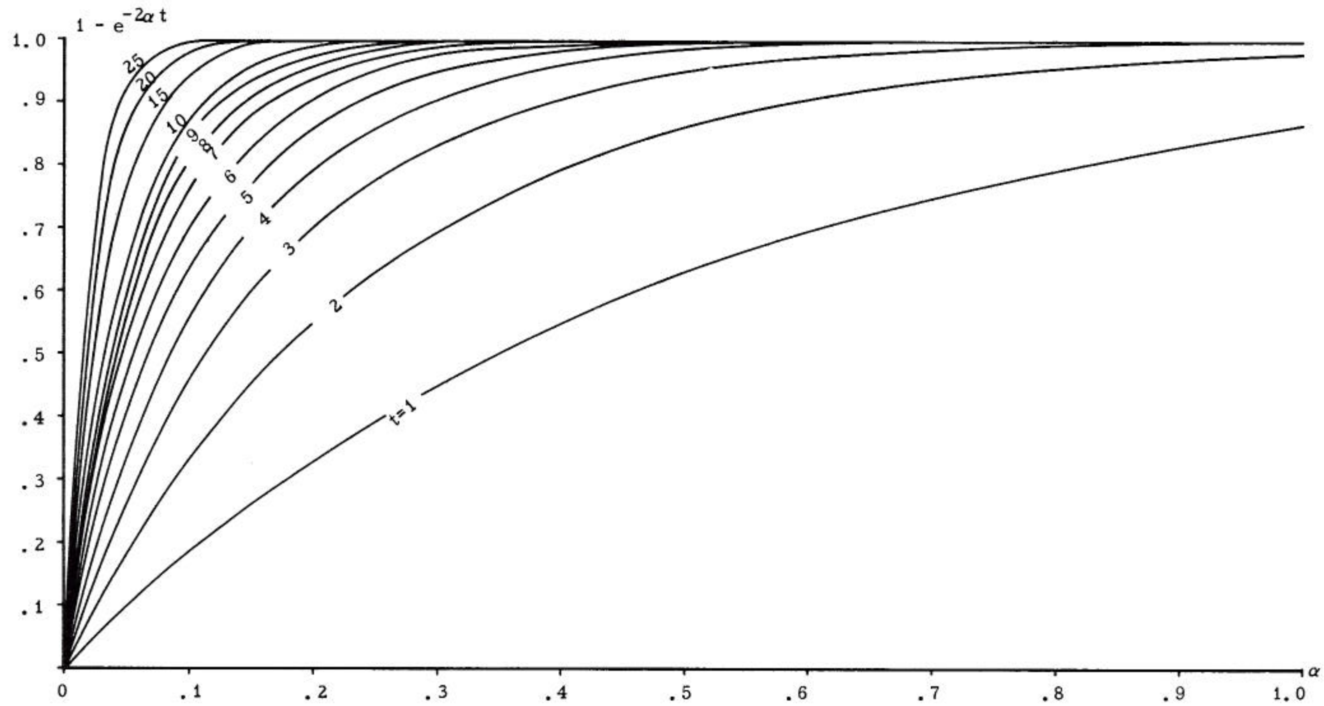

Fig. 2. 2 Expression $1 - e^{-2\alpha t}$ as function of α for various values of t

2. Probability density function of cumulative sum, S. Probability density function of cumulative sum, S, at the time $t = n\Delta t$, are shown in fig. 2. 3 for some particular cases. For the computation of these functions it was assumed that when $t = 0$, $S_o = 0$. The computation was made by using eq. 1. 35 for various values of α, and at different times t.

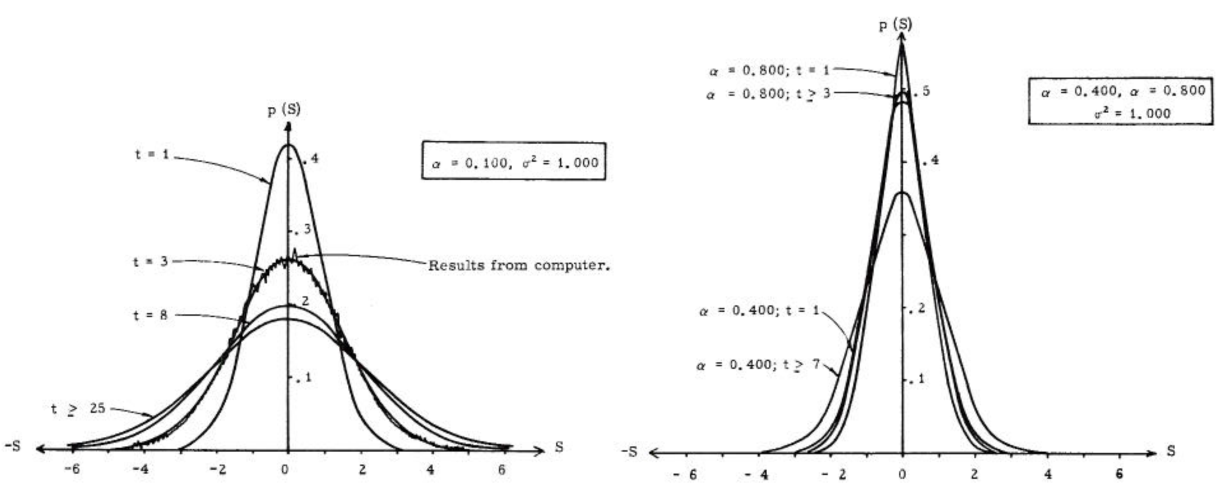

Fig. 2. 3 Probability density functions, p(S), of the cumulative sum, S, at the time t, for various values of α

For $\alpha = 0.100$ and $t = 3$, the frequency density curve obtained by the computer is also given in fig. 2.3. These results are in agreement with the theoretical results as obtained from eq. 1.35. It can be seen from these graphs that for $\alpha = 0.100$ small differences exist between the probability density curves for $t = 8$, and for $t \geq 25$. In fact, the process becomes stationary for $t \geq 23$. For $\alpha = 0.400$, the process becomes stationary for $t \geq 6$. The general conclusion is that the process becomes stationary for approximately $t \geq 2.3/\alpha$. The probability density functions of cumulative sum, S_n, accepted as being unchanged with time t for $t \geq \frac{2.3}{\alpha}$, and they depend only on α for a given σ, or

$$p(S_n, S_o = 0) = \frac{1}{\sigma}\sqrt{\frac{\alpha}{\pi}} \; e^{-\alpha S_n^2/\sigma^2}, \quad -\infty < S < \infty \qquad 2.2$$

where σ^2 is the variance of input. For a normal independent variable with mean zero and variance unity, this equation becomes

$$p(S_n) = \sqrt{\frac{\alpha}{\pi}} \; e^{-\alpha S_n^2}.$$

Figure 2.4 shows the variance of the cumulative sum as a function of α and t. It can be concluded that the variance decreases rapidly with an increase of α, and tends to become constant for a given α and $t \geq \frac{2.3}{\alpha}$, given by

$$\text{Var } S_n = \frac{\sigma^2}{2\alpha} \qquad 2.3$$

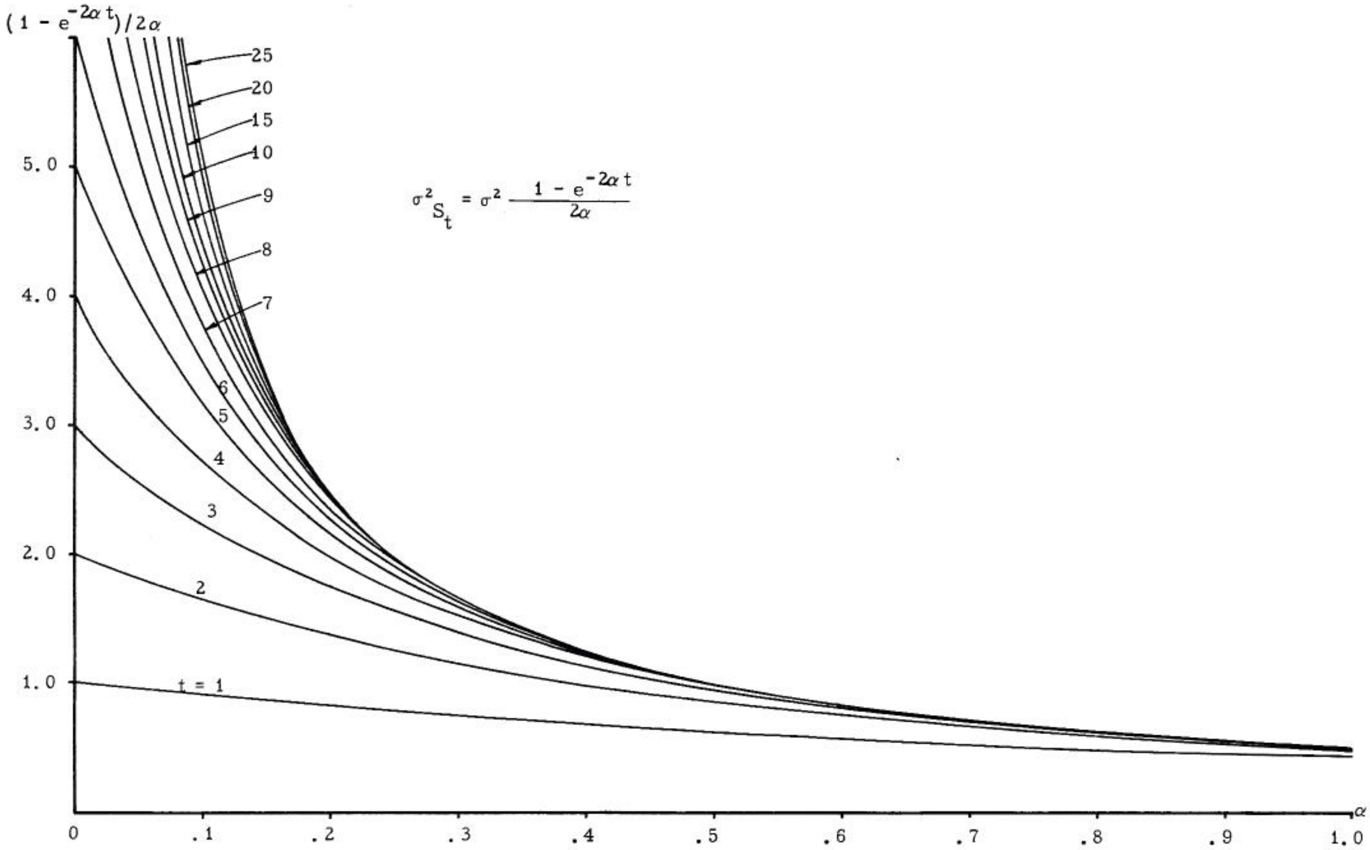

Fig. 2.4 The variance of cumulative sum $S(-\infty < S < \infty)$ as a function of α and t for $\text{Var } x = \sigma^2 = 1$

3. Properties of range. Figure 2.5 shows the mean range for n values between zero and 50, and $\alpha = -0.04, -0.02, 0.00, 0.04, 0.10, 0.20, 0.40, 0.80$ and 2.00. The points represent results obtained by the data generation method and the full lines represent the values of eq. 1.61. The hypothesis tested is that the basic shape of lines is defined by the expression

$$r_n = \sum_{t=1}^{n} \frac{(\text{var } S_t)^{1/2}}{t} \qquad 2.4$$

while the mean range is

$$\bar{R}_n = C_1 \; r_n. \qquad 2.5$$

The constant C_1 is a function only of α. This hypothesis is proved by the results obtained from the computer. Using these results, the expression for the constant C_1 is determined and its values are given by eq. 1.59. It is nearly impossible to calculate the exact values of mean range for $\alpha \neq o$ even for $n = 2$. Calculations obtained on the computer give the same results for the mean range as eq. 1.61, except when $\alpha < o$ and n are large at the

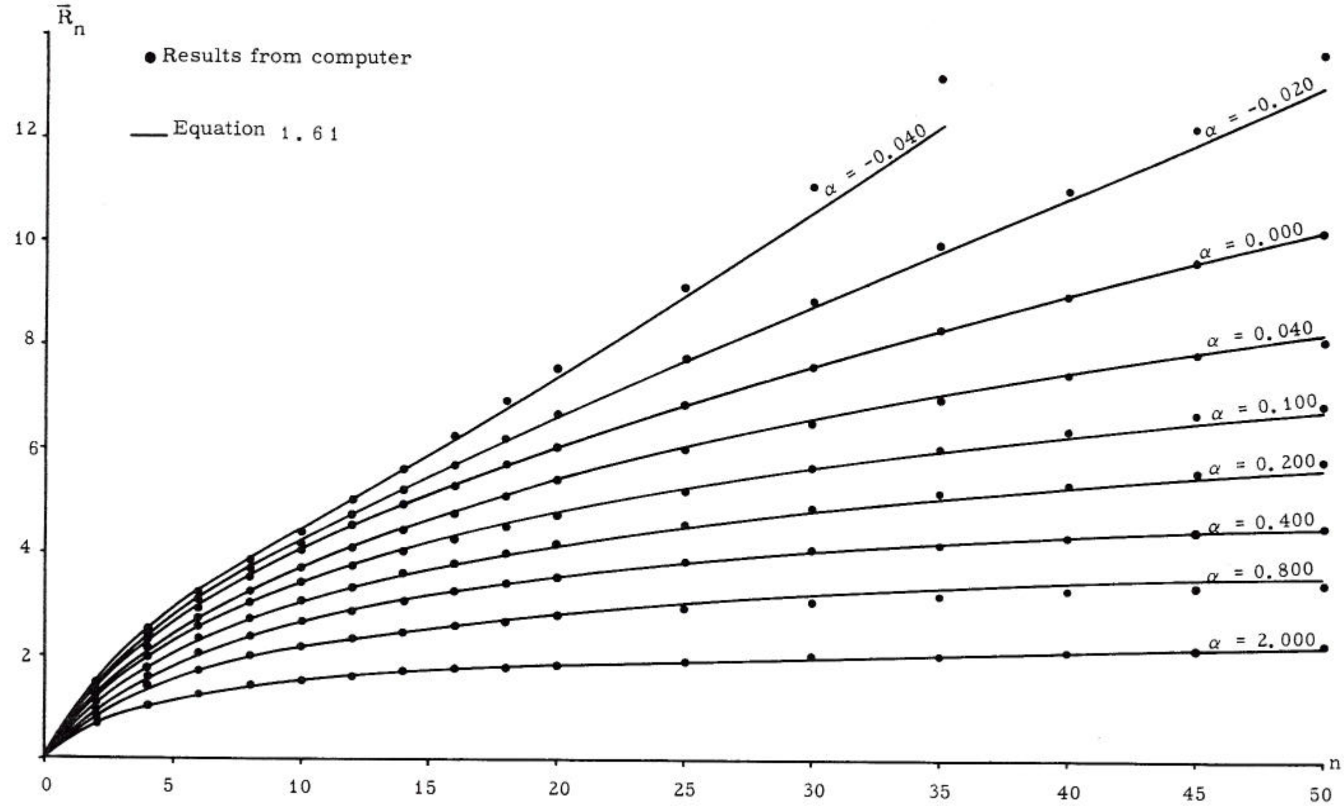

Fig. 2.5 The mean range $\overline{R}_n$ as a function of n and α

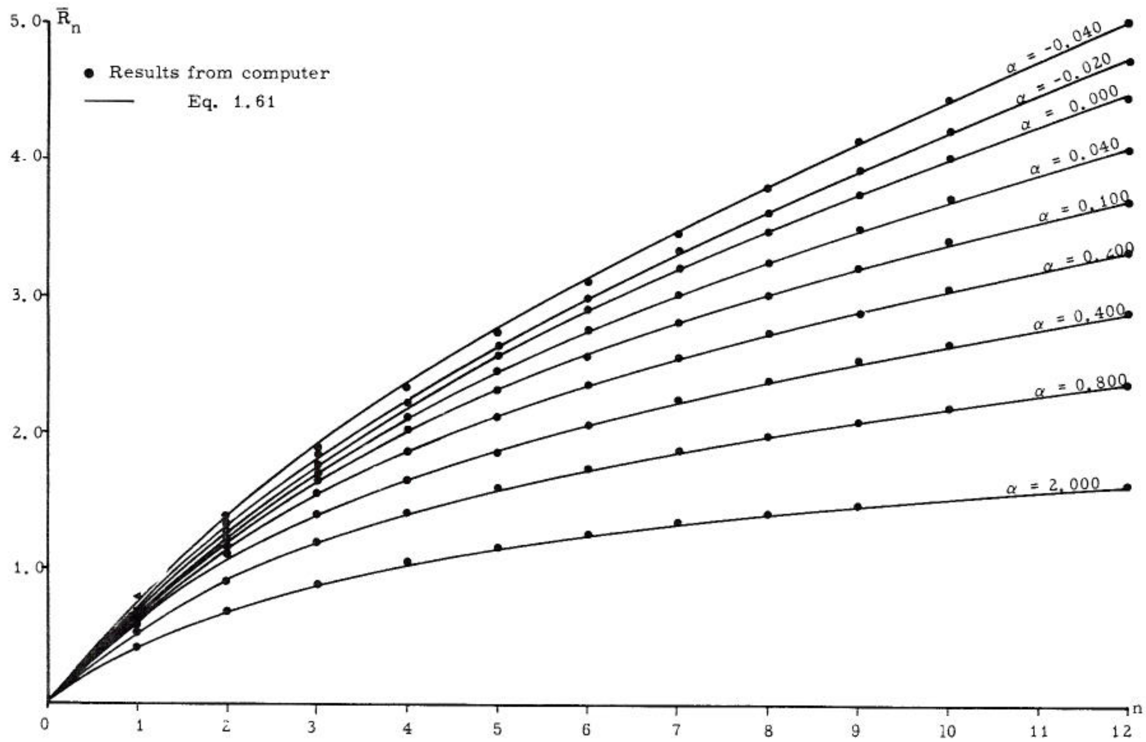

Fig. 2.6 The mean range $\overline{R}_n$ as a function of n and α, for n = 1 - 12, as an enlargement of fig. 2.5 for small n

same moment. When n increases the number of subsamples of size n, as used on the computer, decreases. Therefore, the above differences may be the result of sampling errors. However, for any practical purposes, these differences may be considered as negligible.

The mean range decreases rapidly with an increase of α, but the variance of output increases with α as

$$\mathrm{Var}\ Q_o = 0.5\,\alpha\ (1 - e^{-2\alpha\ t})\ \sigma^2. \qquad 2.6$$

For $\alpha = 2.00$ and sufficiently large values of t, and the variance of output is approximately equal to the variance of input, σ^2.

For $\alpha < 0$, the mean range and the variance of output increases rapidly which is to be expected when the cumulative sum, S, is larger than the output is smaller or vice versa.

Figure 2.7 shows the variance of range for n between zero and 50, and the same α's that have been used for the computation of the mean range. The points represent the results obtained by the data generation method and the lines represent eq. 1.62. The hypothesis tested is that the shape of lines is defined by

$$v_n = \sum_{t=1}^{n} \frac{\mathrm{Var}\ S_t}{t} \qquad 2.7$$

The variance of range is

$$\mathrm{Var}\ R_n = C_2\ v_n. \qquad 2.8$$

The constant C_2 is a function only of α. This hypothesis is proved by the results obtained on the computer. The values determined for the constant C_2 are given by eq. 1.60. The results obtained from

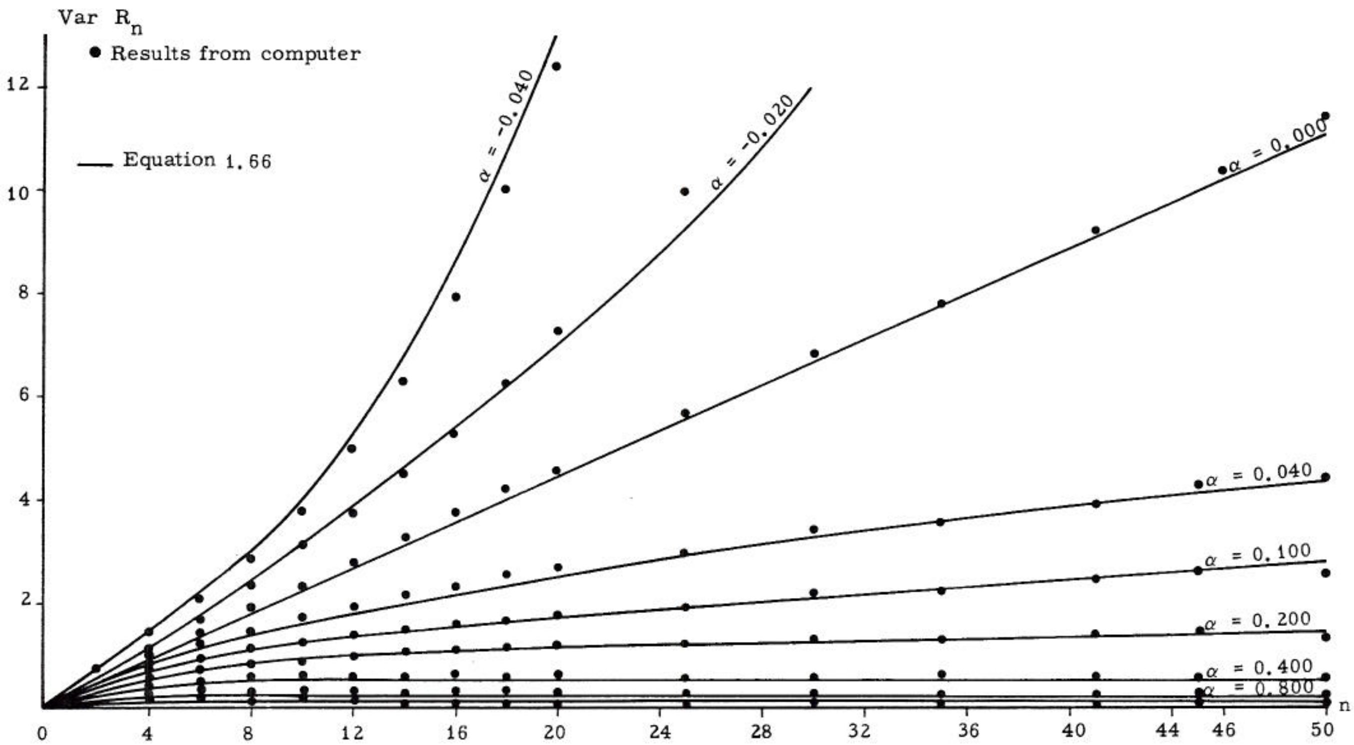

Fig. 2.7 The variance of range, Var R_n as functions of n and α

the computer and the values of eq. 1.66 coincide well except for α smaller than zero. However, even for $\alpha < 0$ the differences are small. The variance of range decreases faster than the mean range with a identical increase of α. The skewness coefficients of range for n between one and 50, and for various values of α as obtained by the data generation method are plotted in fig. 2.9.

The kurtosis of range for various values of n and α as obtained by the data generation method are plotted in fig. 2.10.

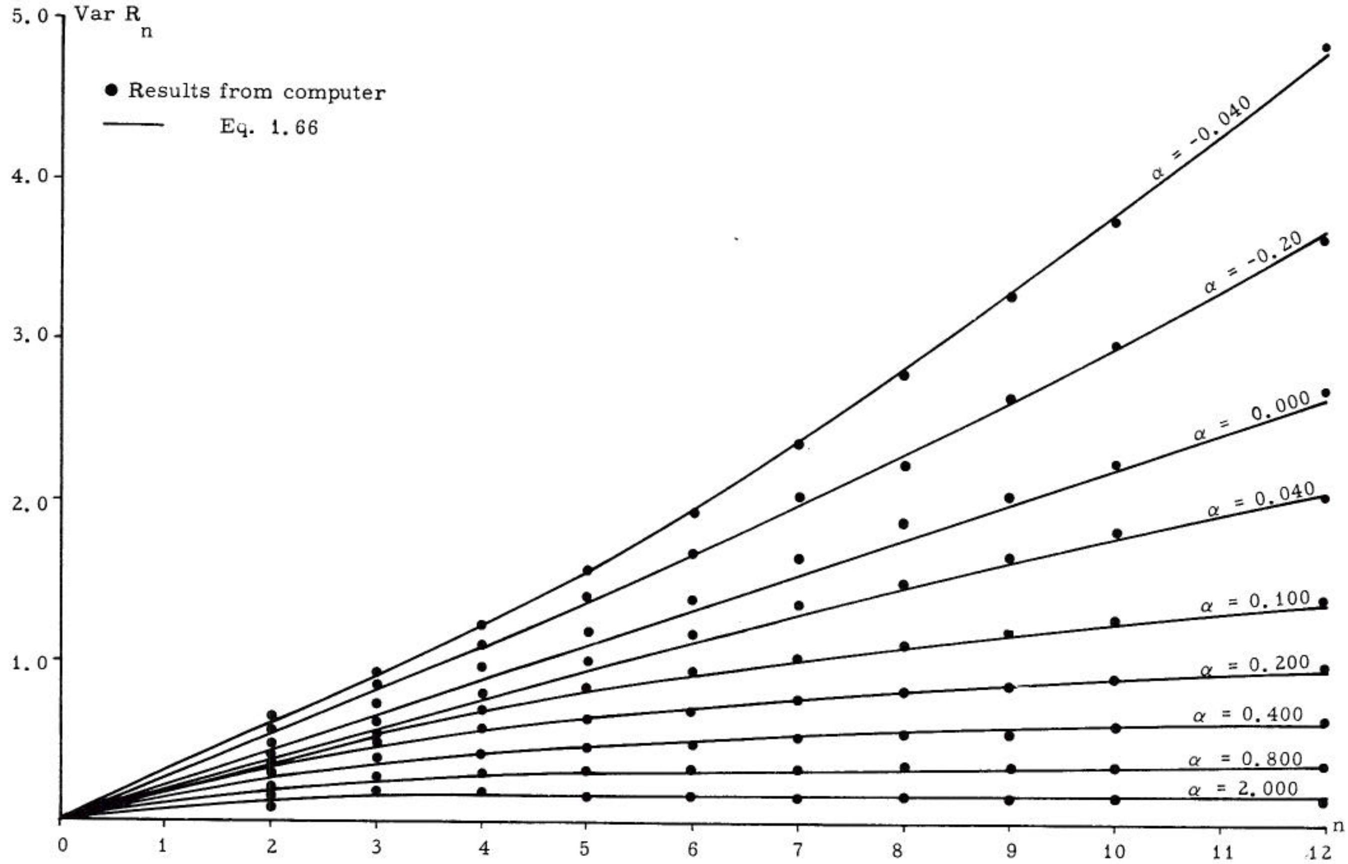

Fig. 2.8 The variance of range, Var R_n as a function of n and α, for n = 1 - 12, as an enlargement of fig. 2.7 for small n

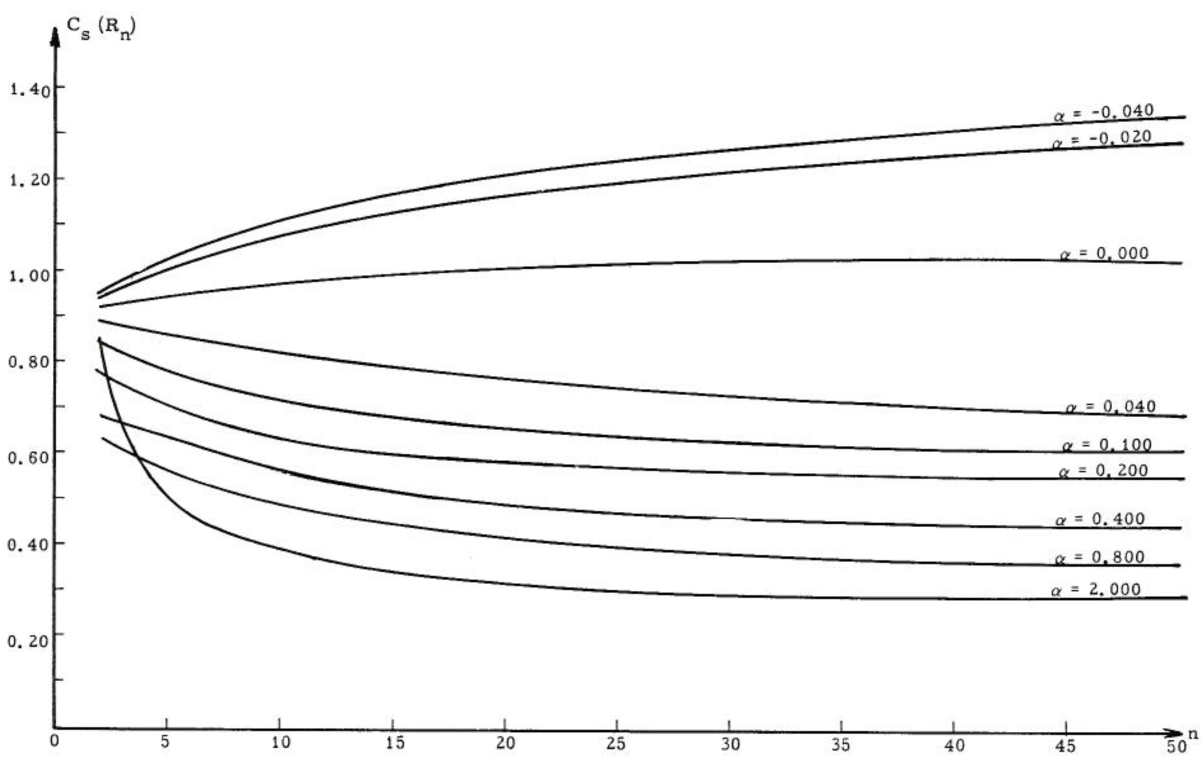

Fig. 2.9 The skewness coefficients of range, $C_s(R_n)$ as a function of n and α

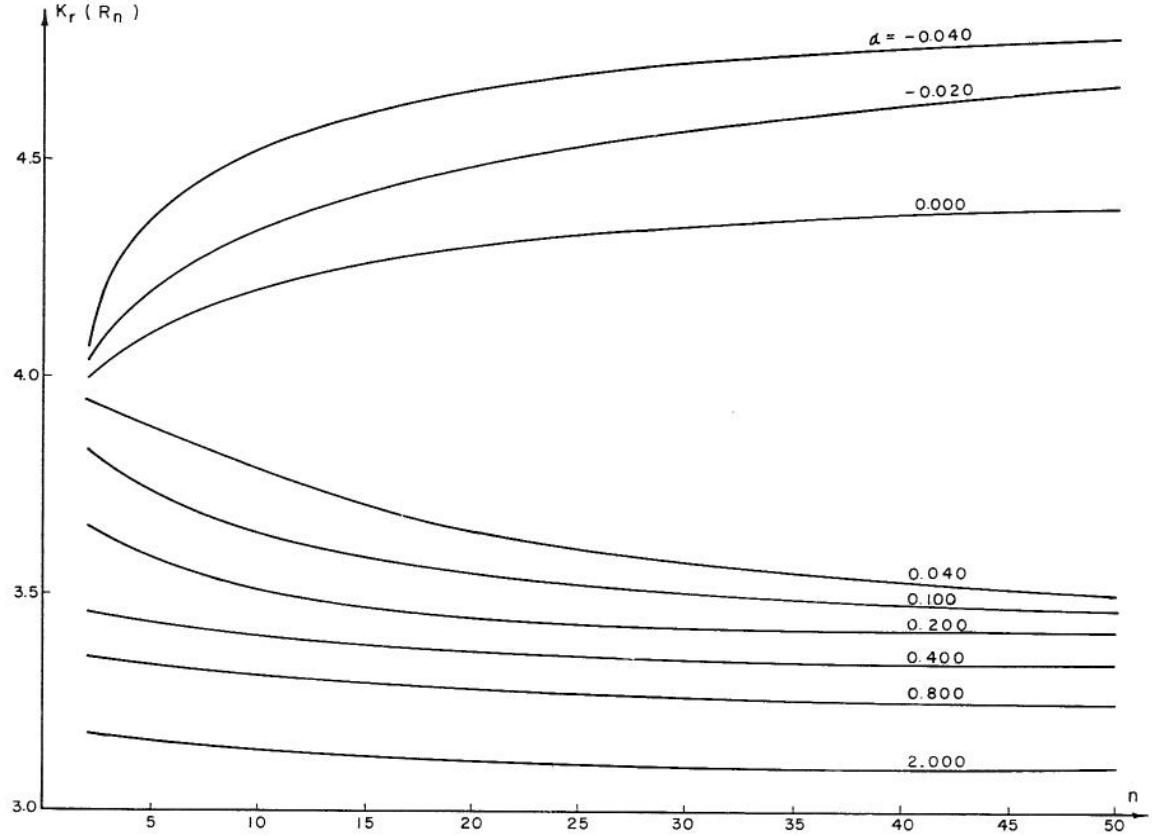

Fig. 2.10 The kurtosis of range, $K_r(R_n)$ as a function of n and α

The mean ranges for n = 5, 10, 25 and 50 are shown in fig. 2.11 as functions of α. It can be concluded from these results that the mean ranges decrease much slower with an increase of α from up to about $\alpha = 0.4$. The mean ranges decrease much slower with an increase of α from 0.4 to 1.0 than for α over 1.00, where the decrease of the mean range with α is very slow.

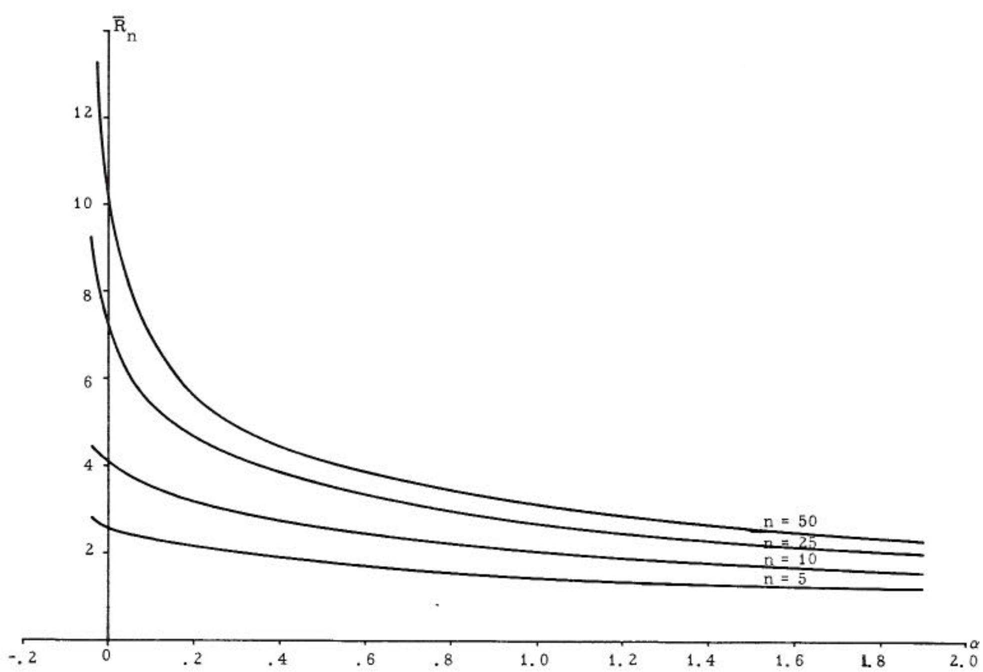

Fig. 2.11 The mean ranges for n = 5, 10, 25 and 50, as functions of α

Probability density functions of range, for a few cases of n, for a given α are plotted in figs. 2.12, 2.13 and 2.14. Results are obtained on the computer. Feller [4] stressed that it is practically impossible to calculate the exact distribution of the range even for n = 3, with simple forms of a distribution of inflows. From figs. 2.12, 2.13 and 2.14, it can be concluded that with an increase of the parameter α the variance of the range decreases. The computer results also show that for $\alpha > 0$ and $n > 5/\alpha$ the probability density functions for the range are approximately normal with the means given in fig. 2.5 and the variances given in fig. 2.7.

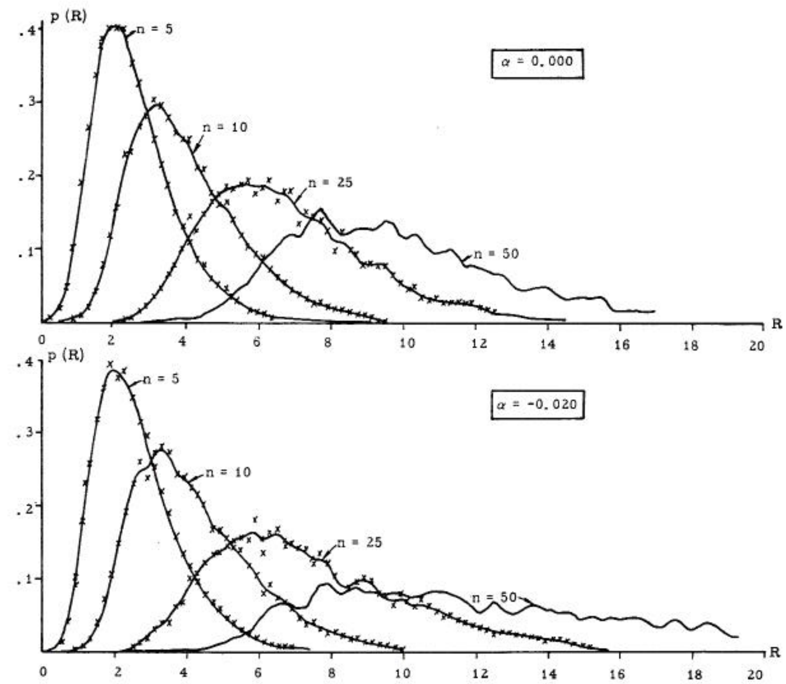

Fig. 2.12 Probability density functions, p(R), of range, for α = 0.00, and α = -0.020, and for n = 5, 10, 25 and 50

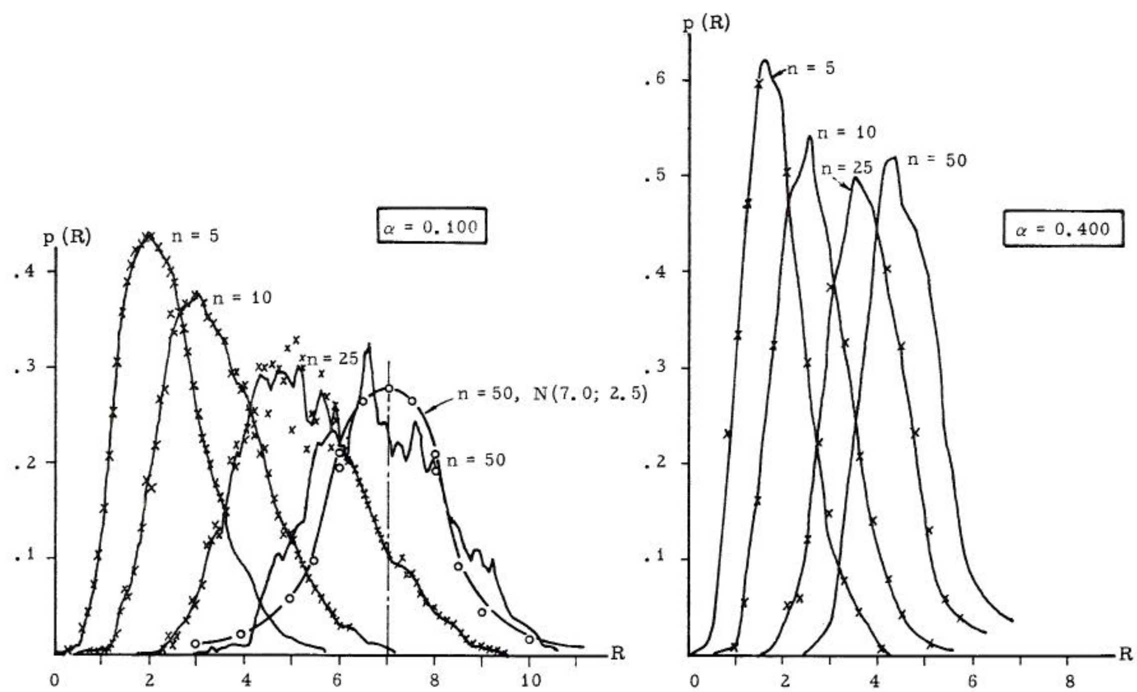

Fig. 2.13 Probability density functions, p(R), of range, for α = 0.100 and α = 0.400, and for n = 5, 10, 26 and 50

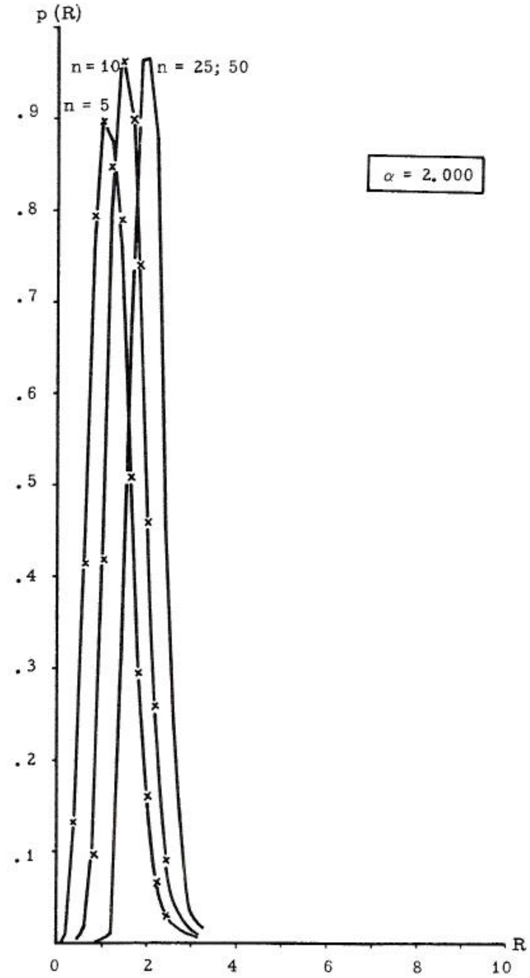

Fig. 2.14 Probability density functions, $p(R)$, of range, for $\alpha = 2.000$ and for $n = 5$, 10 25 and 50

4. Maximum surplus (M_n) and maximum deficit (m_n) of the cumulative sums, S. The mean surplus (M_n) or the mean maximum deficit (m_n) is $E[M_n] = E[m_n] = 1/2\, E[R_n]$. As fig. 2.5 shows the mean range, $E[R_n]$, as a function of n and α, half of the values of range of that graph represents the means of surplus and deficit.

The correlation coefficients $\rho[M_n, m_n]$ between the upper maximum sum (M_n) and the lower minimum sum (m_n) for various values of n are obtained by the data generation method and are plotted in fig. 2.15. For $\alpha = 0$, the correlation coefficient for $n \longrightarrow \infty$ is calculated theoretically by using Feller's variance of range and Anis's variance of the upper maximum sum and lower minimum sum. The same value is obtained on the computer. For $\alpha \neq 0$, the correlation coefficients are obtained only on the computer. For $\alpha = 0$, the correlation coefficient approaches 0.700 with an increase of n, and for $\alpha \neq 0$ it approaches zero.

Figure 2.16 shows the ratio between the variance of range (Var R_n) and the variance of maximum surplus (Var M_n) or maximum deficit (Var m_n). The variance of maximum surplus or maximum deficit is defined by

$$\mathrm{Var}\,[M_n] = \mathrm{Var}\,[m_n] = \frac{\mathrm{Var}\,[R_n]}{2(1-\rho\,[M_n, m_n])}. \qquad 2.9$$

The skewness coefficients of maximum surplus (M_n) or maximum deficit (m_n) for n between two and 50 and for various values of α are plotted in fig. 2.17. For $0 \leq \alpha \leq 0.200$ the skewness coefficients decrease with an increase of n. For $\alpha < 0$ and $\alpha > 0.200$ the skewness coefficients decrease with an increase of n only for small values of n, and increase for large values of n. These results are obtained on the computer by the data generation method.

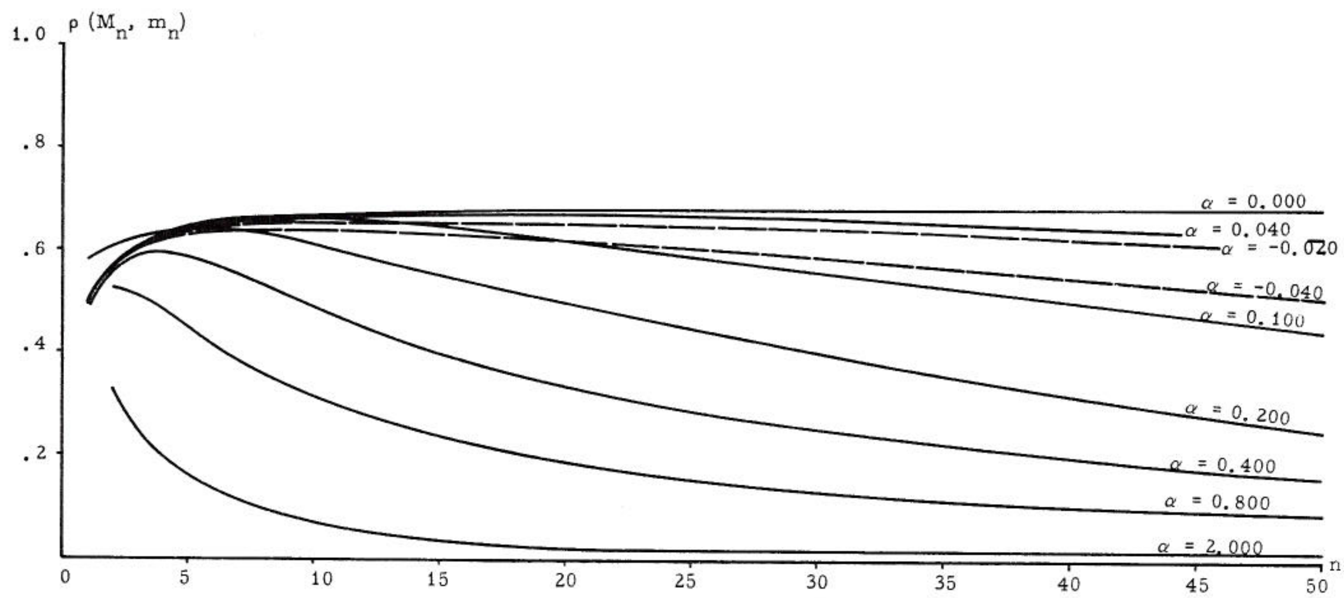

Fig. 2.15 Correlation coefficients between M_n and m_n, for various values of α as function of n

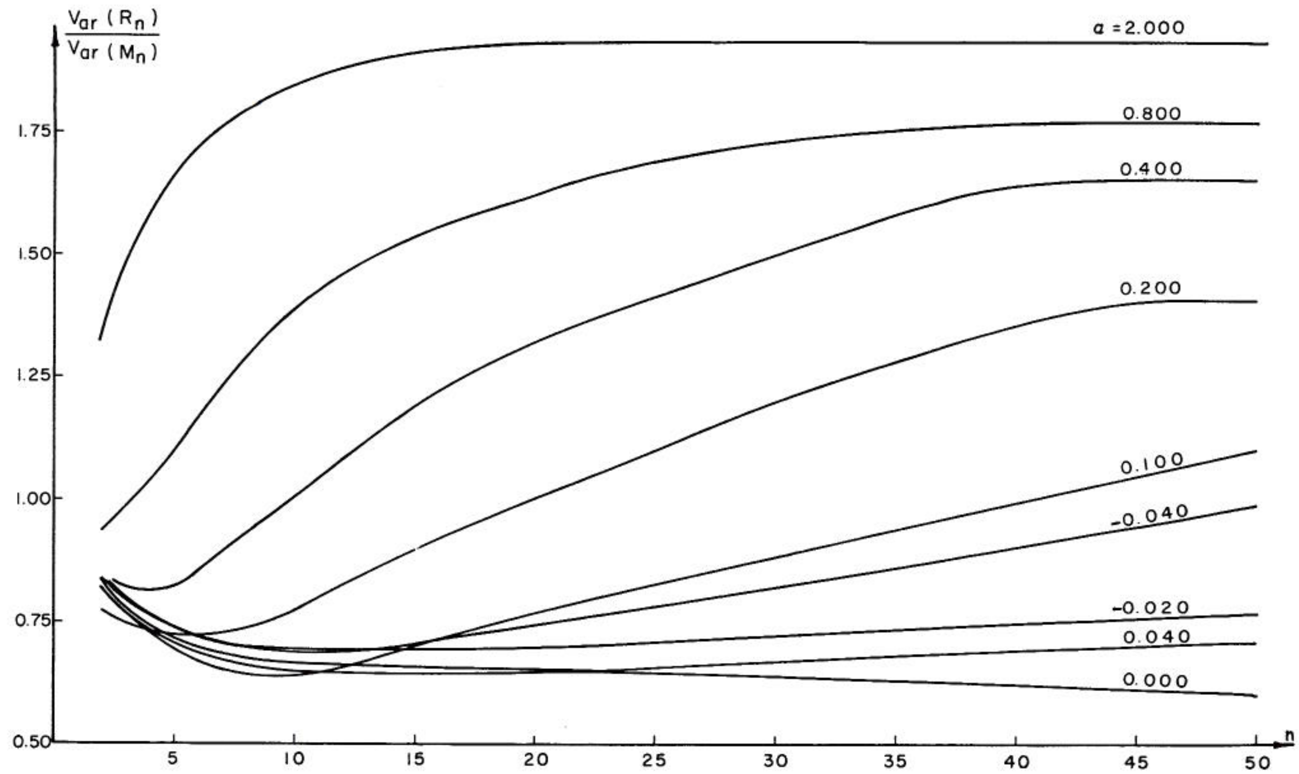

Fig. 2.16 The ratio between Var $[R_n]$ and Var $[M_n]$ with (Var $[m_n]$ = Var $[M_n]$

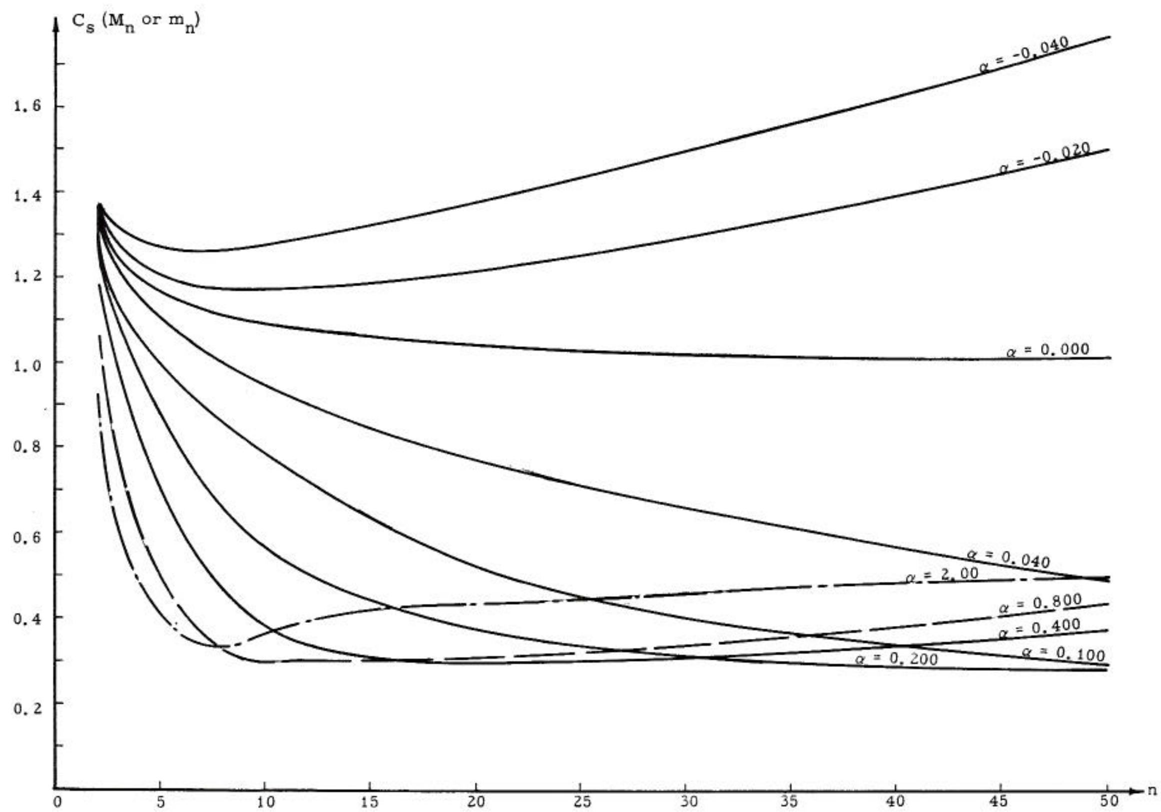

Fig. 2.17 Skewness coefficients of maximum surplus (M_n) or maximum deficit (m_n) as a function of n and α

Figure 2.18 shows the kurtosis of maximum surplus (M_n) or maximum deficit (m_n) obtained on the computer by the data generation method. It clearly shows how the kurtosis changes as a function of α and n.

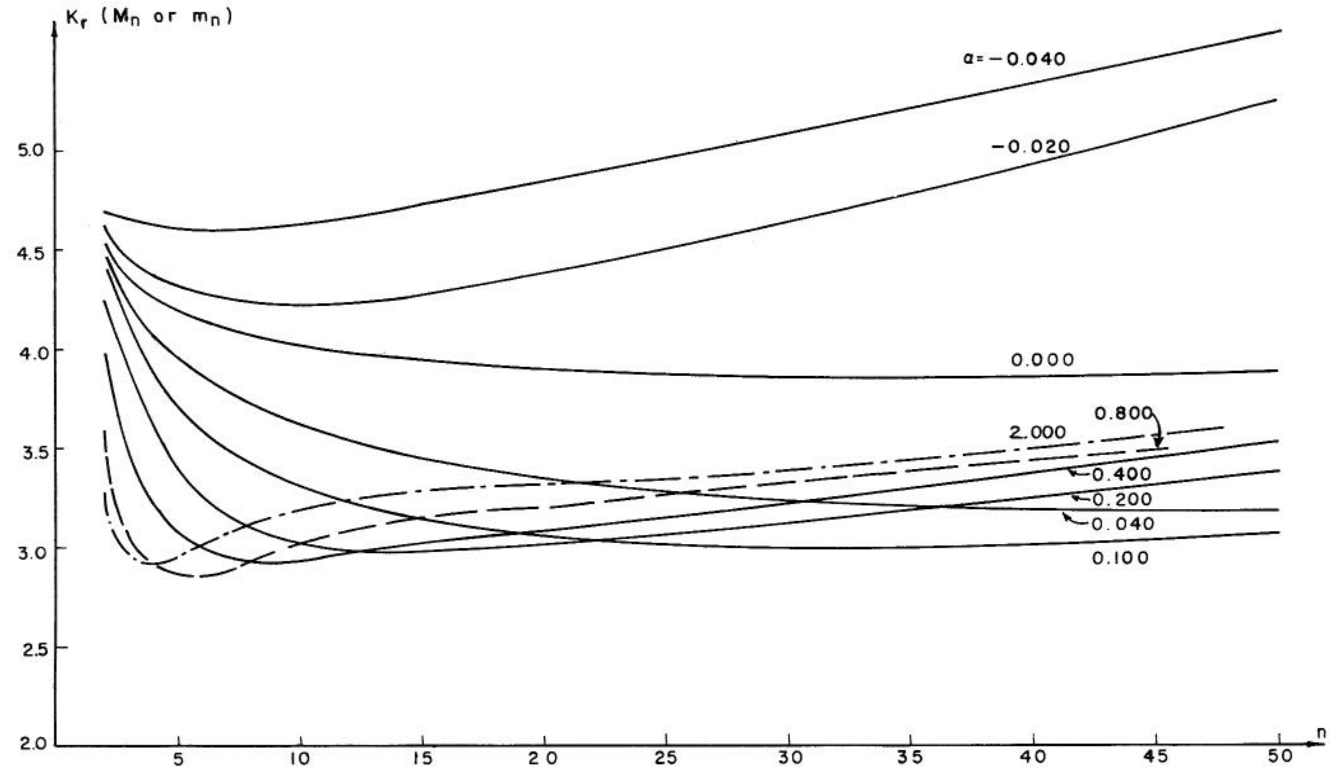

Fig. 2.18 The kurtosis of maximum surplus (M_n) or maximum deficit (m_n) as a function of n and α

The discrete probability or probability mass for $M_n = 0$ or $m_n = 0$ for $1 \leq n \leq 50$ is shown in fig. 2.19 as a function of α and n. It can be concluded from these results that the probability mass at $M_n = 0$ (or $m_n = 0$) decreases rapidly with an increase of both α and n. Results are obtained on the computer by the data generation method. Some probability density functions of M_n (or m_n) for several values of n are plotted in figs. 2.20, 2.21 and 2.22 for various values of α. It can be concluded from these figures that the variance of the maximum surplus M_n decreases with an increase of the parameter α. In fig. 2.21 for n = 50 and α = 0.100, it is shown that the probability density function of M_n is approximately normal with the mean 3.50 and variance 2.30. The probability mass for $M_n = 0$ is not shown in figs. 2.20, 2.21 and 2.22. The probability density functions of M_n are plotted in fig. 2.20 as smooth curves. The smooth curves were used to show the differences between these curves and those developed for various values of n. Figures 2.21 and 2.22 show the curves as they came from the computer, obtained by the data generation method.

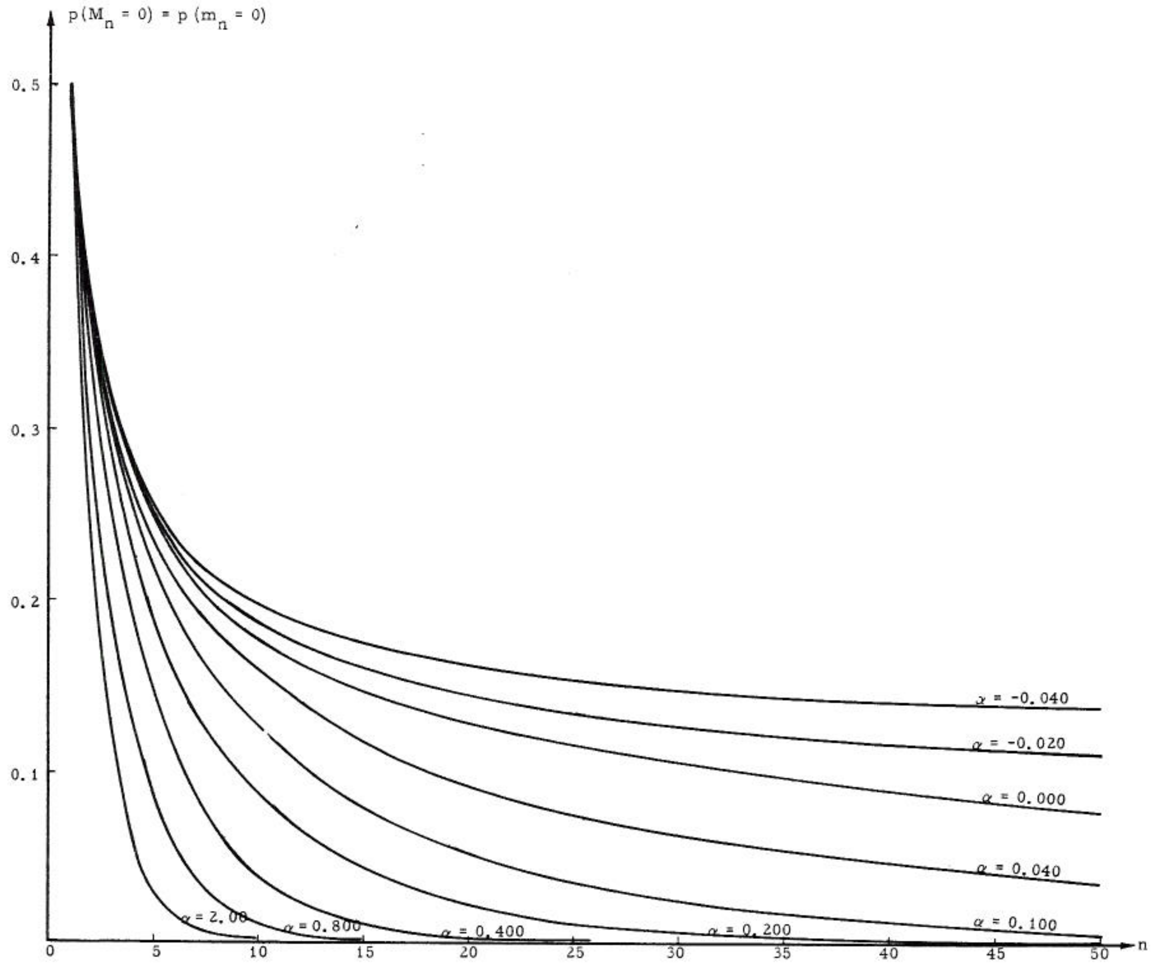

Fig. 2.19 Discrete probability or probability mass, $p\,(M_n = 0)$ or $p\,(m_n = 0)$ at zero values, $M_n = 0$ or $m_n = 0$, for $-0.04 \leq \alpha \leq 2.00$, and $1 \leq n \leq 50$

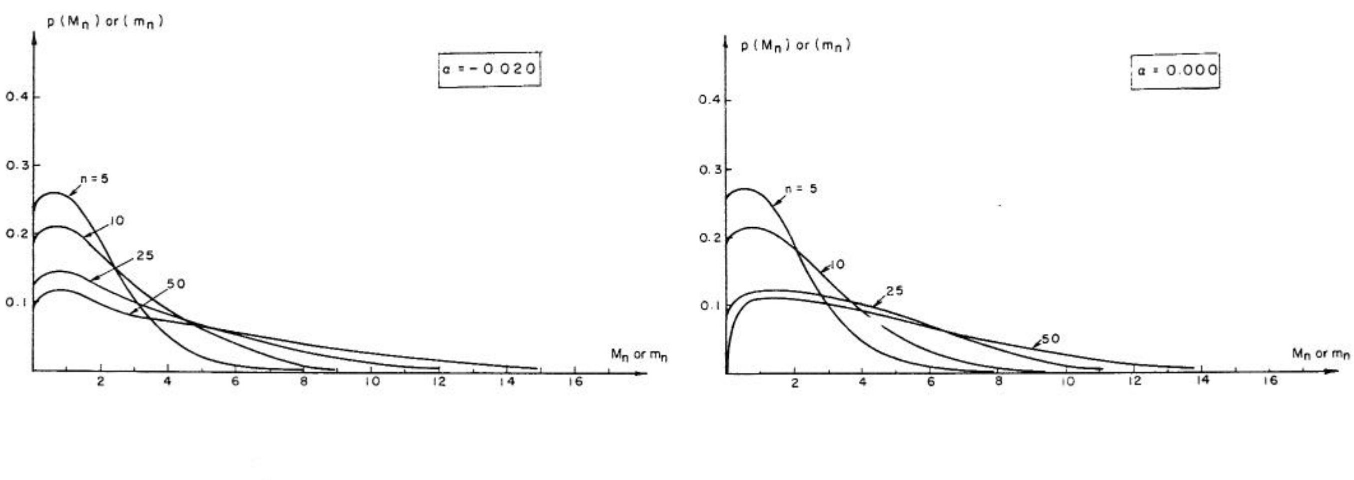

Fig. 2.20 Probability density curves, $p\,(M_n)$ or $p\,(m_n)$, for $\alpha = 0.00$ and $\alpha = 0.020$, and for n = 5, 10, 25 and 50. The probability mass for $M_n = 0$ or $m_n = 0$ is not shown on these graphs

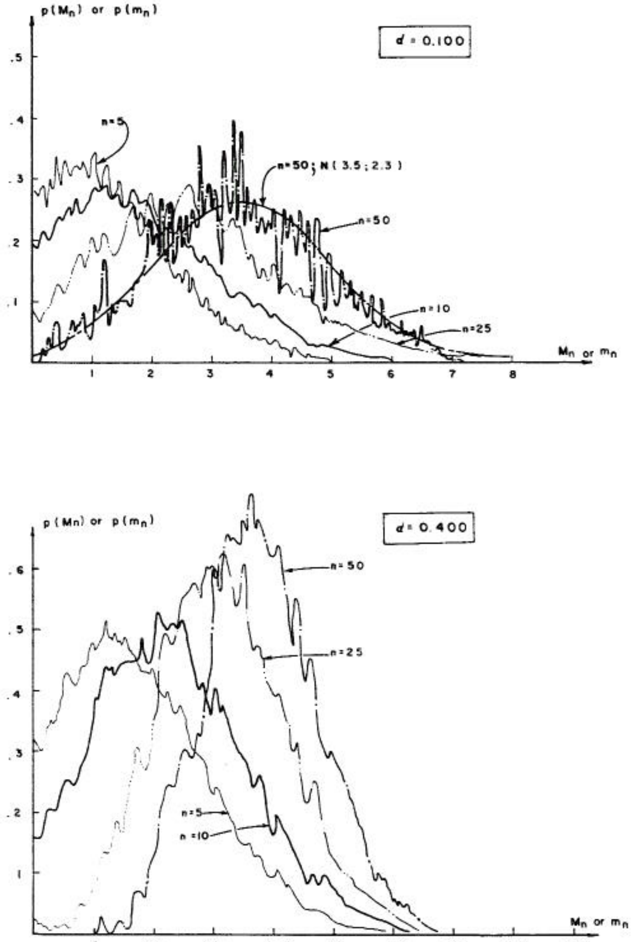

Fig. 2.21 Probability density curves, p (M_n) or p (m_n), for $\alpha = 0.100$ and $\alpha = 0.400$ and for n = 5, 10, 25 and 50. The probability mass for $M_n = 0$ or $m_n = 0$ is not shown on these graphs

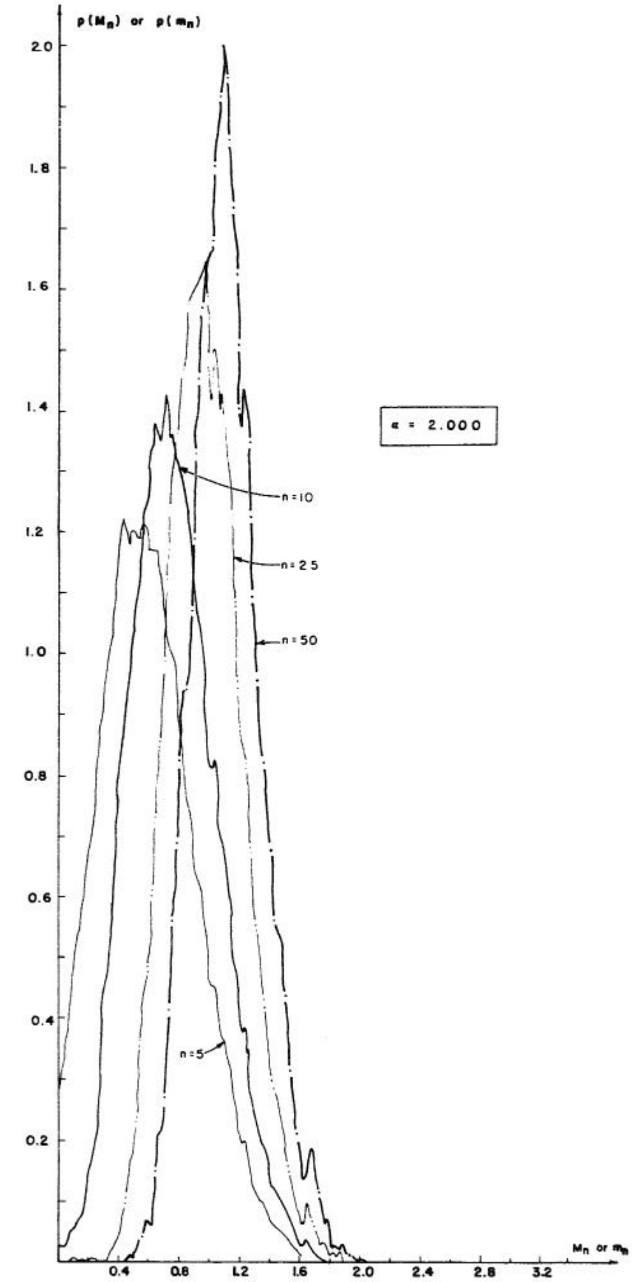

Fig. 2.22 Probability density curves, p (M_n) or p (m_n), for $\alpha = 2.000$ and for n = 5, 10, 25 and 50. The probability mass for $M_n = 0$ or $m_n = 0$ is not shown on these graphs

CHAPTER III

DISTRIBUTIONS OF RANGE, SURPLUS AND DEFICIT AS OBTAINED ON THE DIGITAL COMPUTER BY THE DATA GENERATION METHOD

When the output is a linear function of the cumulative sum it is useful to know the distributions of range, surplus and deficit for the normal independent variable of the input. These distributions are computed by the data generation method, employing a CDC 3600 digital computer, from 100,000 normal independent random numbers. For different values of α the distributions are presented for the range and the surplus for various values of n. As the surplus is equal to the deficit, only the distributions of surplus are presented in this chapter.

The mean range is given in table 3.1 for n = 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 45 and 50 for α = -0.04, -0.02, 0.00, 0.04, 0.10, 0.20, 0.40, 0.80 and 2.00.

The variance of range is given in table 3.2. The correlation coefficient between the maximum surplus (M_n) and the maximum deficit (m_n) are given in table 3.3 for each of the above parameters of α and n.

From the data given in table 3.1, the mean maximum surplus (M_n) and the mean maximum deficit (m_n) can be computed by using the equation $E[M_n] = E[m_n] = 1/2\,E[R_n]$.

By using the data from tables 3.2 and 3.3, and eq. 2.9, the variance for maximum surplus (M_n) or for maximum deficit (m_n) can be computed.

Distribution functions of range for α = -0.040, and n = 2, 4, 6, 8, 12, 16, 20, 25 and 30 are plotted in fig. 3.1. Distribution functions of range for α = -0.020, 0.000, 0.040, 0.100, 0.200, 0.400, 0.800, and 2.000, and for n = 2, 4, 6, 8, 12, 16, 20, 30, 40 and 50 are plotted in figs. 3.2 through 3.9.

For some values of α and n, distribution functions of maximum surplus (M_n) or of maximum deficit (m_n) are plotted in figs. 3.10 through 3.18.

TABLE 3.1

Ratio of mean of range to standard deviation of the input, $\bar{R}_n/\sigma$, obtained on the digital computer

n/α	-0.040	-0.020	0.000	0.040	0.100	0.200	0.400	0.800	2.000
2	1.402	1.382	1.363	1.325	1.274	1.200	1.082	0.923	0.680
3	1.894	1.858	1.823	1.760	1.678	1.564	1.398	1.191	0.887
4	2.327	2.271	2.221	2.129	2.013	1.862	1.654	1.410	1.045
5	2.727	2.650	2.581	2.458	2.308	2.120	1.876	1.595	1.168
6	3.096	2.995	2.904	2.749	2.566	2.346	2.070	1.751	1.267
7	3.459	3.328	3.213	3.021	2.802	2.551	2.242	1.883	1.347
8	3.784	3.623	3.483	3.257	3.009	2.733	2.397	2.003	1.421
9	4.134	3.931	3.760	3.490	3.204	2.898	2.534	2.103	1.478
10	4.461	4.221	4.022	3.714	3.398	3.065	2.666	2.195	1.533
12	5.065	4.737	4.477	4.094	3.719	3.340	2.889	2.354	1.623
14	5.682	5.246	4.917	4.452	4.025	3.606	3.096	2.489	1.696
16	6.312	5.746	5.332	4.780	4.292	3.824	3.259	2.597	1.756
18	6.934	6.217	5.716	5.079	4.545	4.032	3.406	2.691	1.808
20	7.595	6.687	6.080	5.348	4.764	4.205	3.523	2.763	1.851
25	9.269	7.787	6.898	5.949	5.264	4.610	3.807	2.945	1.948
30	11.152	8.897	7.673	6.513	5.711	4.936	4.018	3.078	2.023
35	13.372	9.951	8.320	6.947	6.071	5.218	4.204	3.195	2.086
40	15.784	11.042	9.007	7.428	6.430	5.457	4.348	3.281	2.137
45	19.000	12.267	9.642	7.814	6.718	5.648	4.477	3.365	2.183
50	23.402	13.636	10.191	8.145	6.984	5.834	4.589	3.435	2.221

TABLE 3.2

Ratio of variance of range to the variance of the input, $\sigma^2 [R_n] / \sigma^2$, obtained on the digital computer

n/α	-0.040	-0.020	0.000	0.040	0.100	0.200	0.400	0.800	2.000
2	0.648	0.623	0.600	0.556	0.500	0.427	0.331	0.236	0.148
3	0.923	0.869	0.819	0.734	0.633	0.514	0.382	0.275	0.172
4	1.232	1.135	1.050	0.910	0.755	0.590	0.429	0.308	0.179
5	1.558	1.404	1.273	1.067	0.853	0.647	0.464	0.327	0.177
6	1.912	1.687	1.501	1.221	0.950	0.709	0.506	0.346	0.173
7	2.370	2.040	1.777	1.398	1.054	0.767	0.536	0.349	0.168
8	2.683	2.259	1.934	1.485	1.106	0.810	0.567	0.358	0.163
9	3.288	2.689	2.247	1.664	1.197	0.858	0.588	0.360	0.160
10	3.716	2.970	2.439	1.766	1.256	0.896	0.607	0.357	0.153
12	4.849	3.682	2.910	2.012	1.397	0.990	0.645	0.357	0.146
14	6.194	4.439	3.360	2.209	1.503	1.060	0.664	0.350	0.140
16	7.767	5.269	3.817	2.386	1.587	1.104	0.671	0.344	0.133
18	9.935	6.280	4.320	2.585	1.715	1.169	0.681	0.338	0.131
20	12.359	7.331	4.812	2.759	1.807	1.210	0.682	0.334	0.128
25	20.291	10.001	5.790	3.003	1.931	1.257	0.675	0.320	0.121
30	32.423	13.496	6.982	3.453	2.220	1.342	0.668	0.304	0.114
35	51.970	17.549	7.798	3.530	2.260	1.346	0.655	0.296	0.107
40	86.768	24.183	9.426	3.971	2.438	1.417	0.675	0.298	0.105
45	134.872	31.127	10.686	4.387	2.631	1.417	0.635	0.280	0.101
50	211.738	39.442	11.550	4.404	2.540	1.356	0.620	0.267	0.096

TABLE 3.3

Correlation coefficient, $\rho (M_n, m_n)$, between the upper maximum sum, M_n, and the lower minimum sum, m_n

n/α	-0.040	-0.020	0.000	0.040	0.100	0.200	0.400	0.800	2.000
2	0.567	0.570	0.570	0.575	0.575	0.615	0.568	0.528	0.330
3	0.605	0.607	0.615	0.612	0.616	0.614	0.595	0.516	0.250
4	0.621	0.626	0.636	0.634	0.635	0.630	0.596	0.480	0.195
5	0.633	0.636	0.642	0.646	0.650	0.640	0.580	0.450	0.160
6	0.633	0.641	0.648	0.654	0.657	0.640	0.570	0.410	0.135
7	0.635	0.654	0.650	0.657	0.658	0.640	0.553	0.392	0.118
8	0.645	0.650	0.664	0.667	0.662	0.630	0.526	0.356	0.095
9	0.640	0.654	0.660	0.669	0.666	0.627	0.511	0.335	0.080
10	0.640	0.656	0.665	0.670	0.664	0.614	0.494	0.315	0.080
12	0.640	0.657	0.662	0.670	0.654	0.586	0.446	0.272	0.060
14	0.635	0.656	0.670	0.673	0.648	0.560	0.410	0.247	0.040
16	0.635	0.655	0.670	0.676	0.641	0.544	0.380	0.222	0.040
18	0.632	0.654	0.672	0.675	0.630	0.518	0.360	0.207	0.035
20	0.617	0.662	0.674	0.678	0.632	0.500	0.342	0.200	0.025
25	0.603	0.646	0.682	0.680	0.605	0.450	0.285	0.156	0.005
30	0.588	0.646	0.680	0.665	0.550	0.400	0.260	0.146	0.020
35	0.573	0.646	0.694	0.662	0.548	0.375	0.235	0.134	0.030
40	0.536	0.620	0.678	0.648	0.500	0.320	0.181	0.096	0.020
45	0.522	0.610	0.675	0.622	0.455	0.300	0.190	0.100	0.020
50	0.522	0.614	0.692	0.635	0.455	0.300	0.185	0.108	0.040

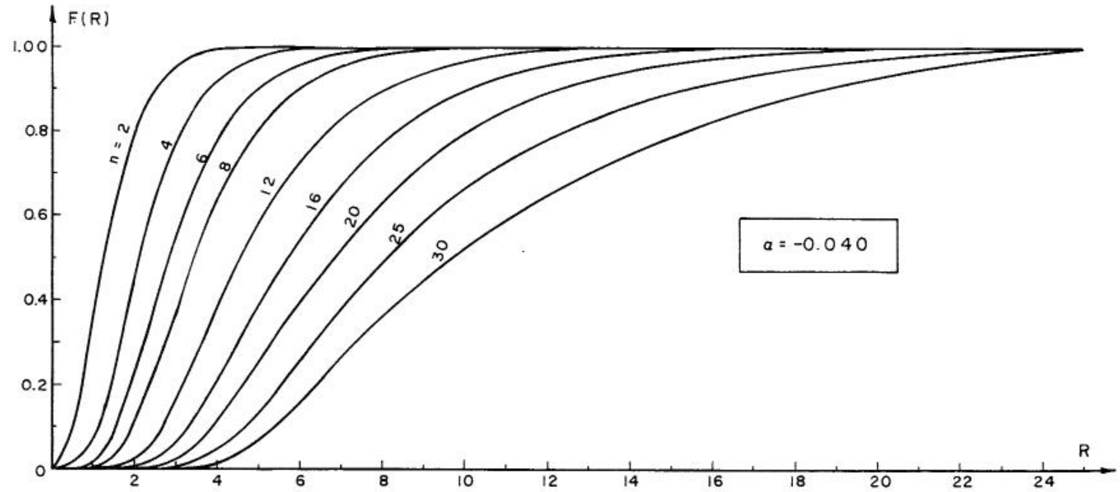

Fig. 3.1 Distribution functions of range, F (R), for α = -0.040 and for various values of n

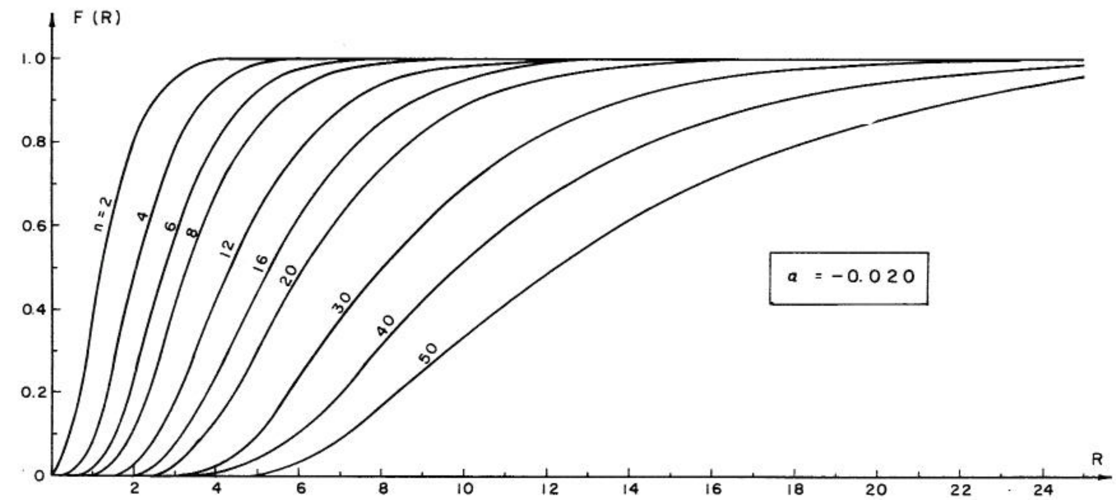

Fig. 3.2 Distribution functions of range, F (R), for α = -0.020 and various values of n

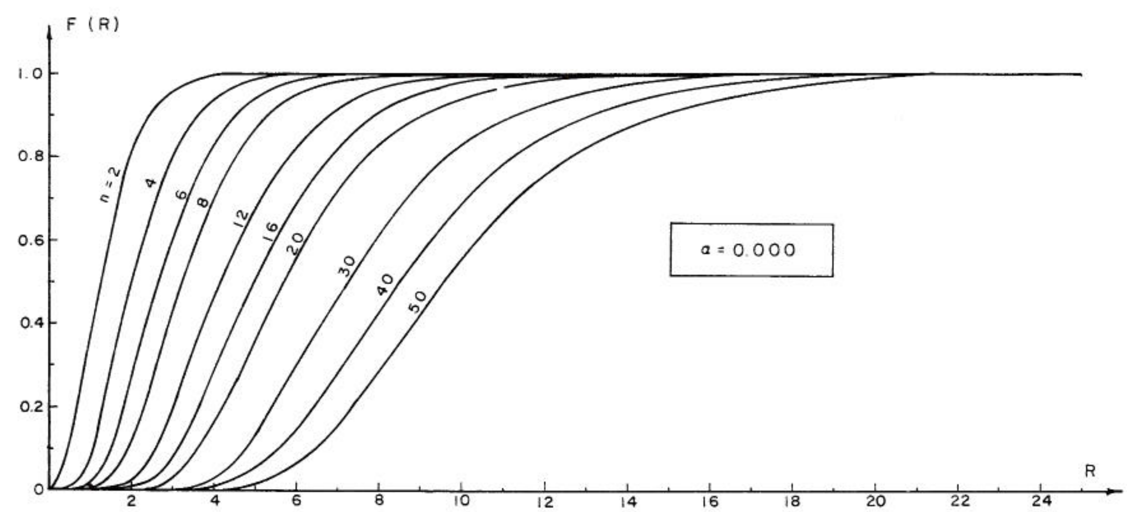

Fig. 3.3 Distribution functions of range, F (R), for α = 0.000 and various values of n

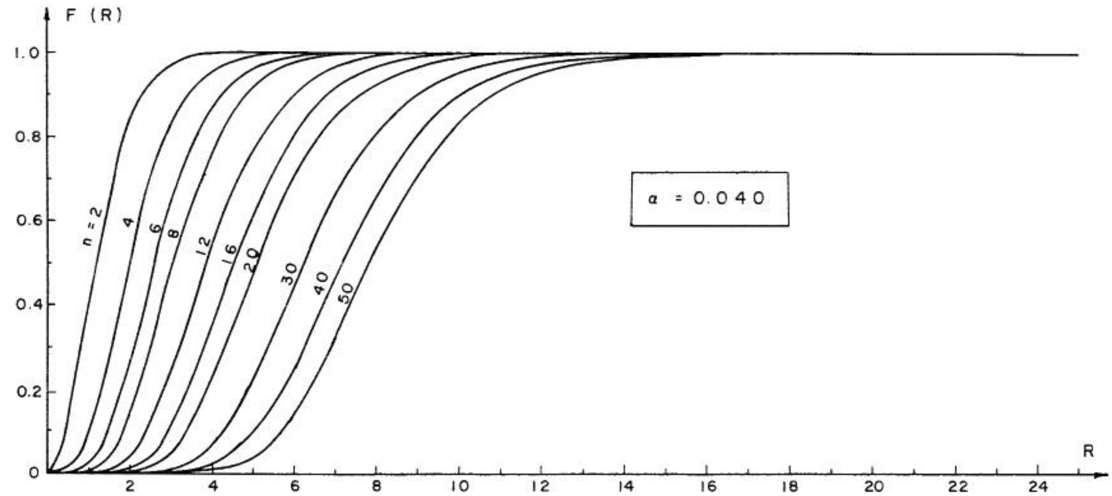

Fig. 3.4 Distribution functions of range, F (R), for α = 0.040 and various values of n

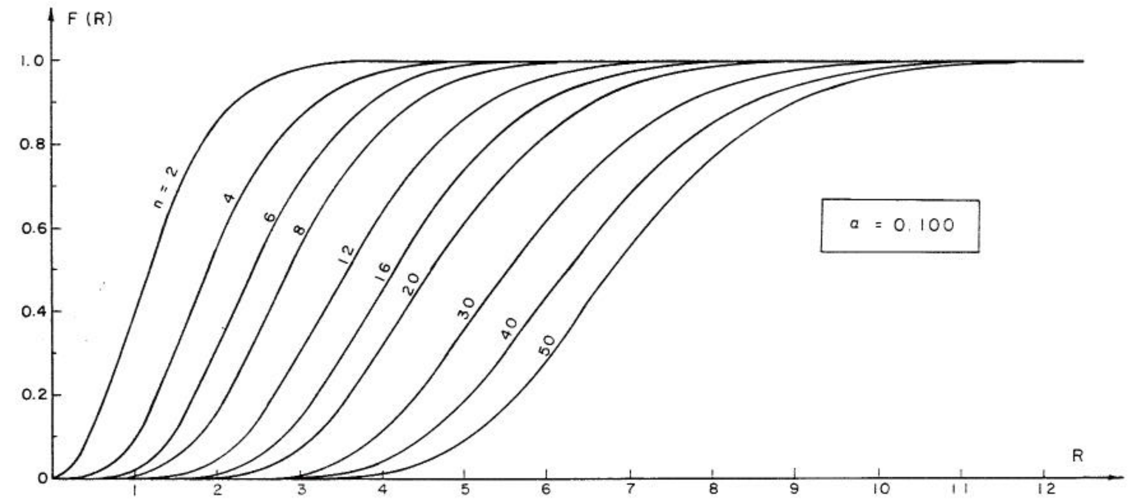

Fig. 3.5 Distribution functions of range, F (R), for α = 0.100 and various values of n

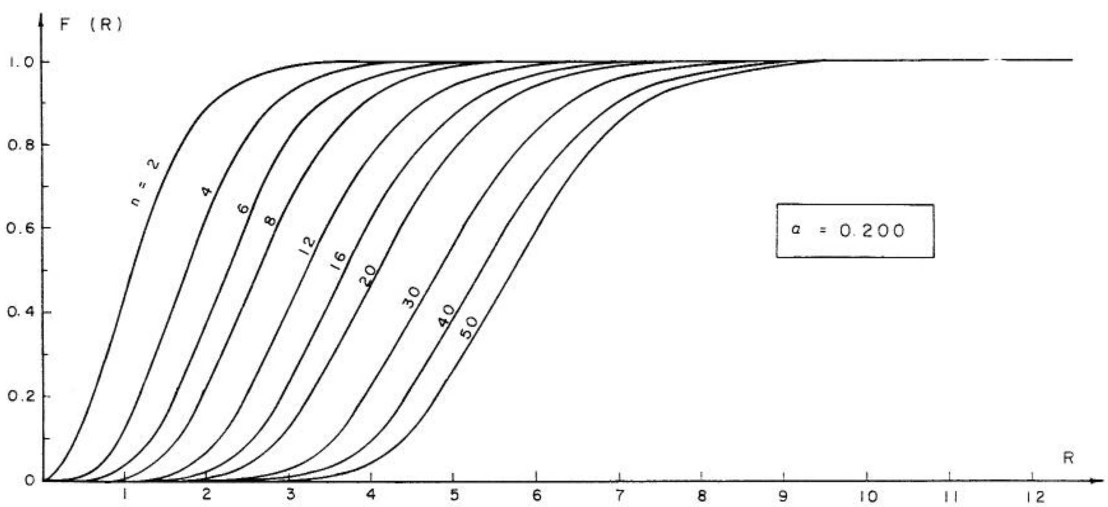

Fig. 3.6 Distribution functions of range, F (R), for α = 0.200 and various values of n

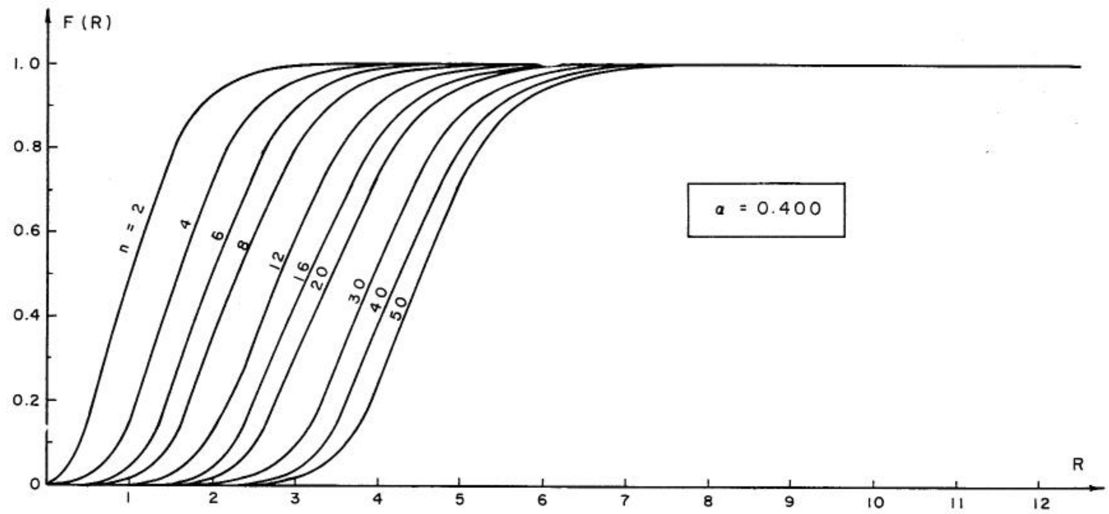

Fig. 3.7 Distribution functions of range, F (R), for α = 0.400 and various values of n

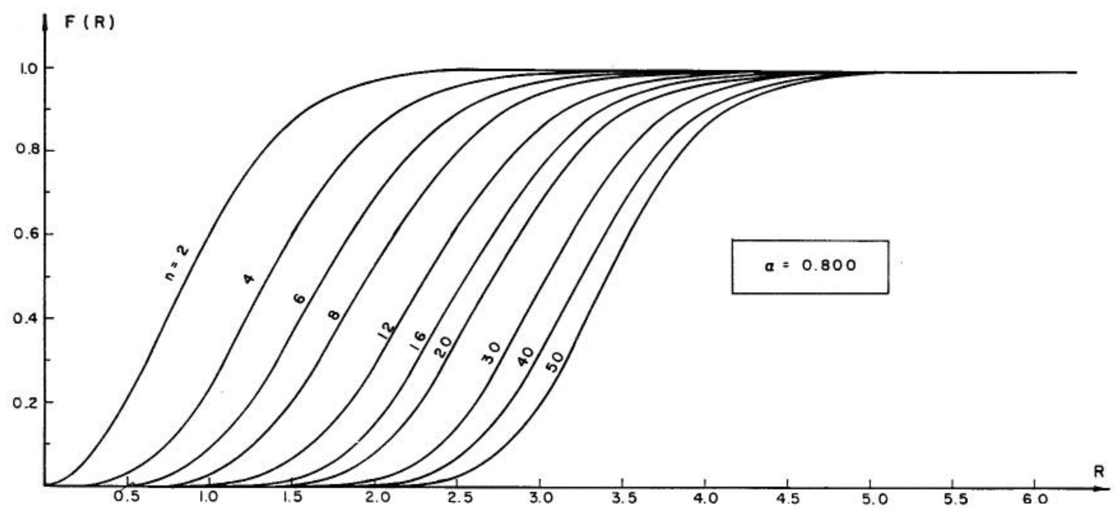

Fig. 3.8 Distribution functions of range, F (R), for α = 0.800 and various values of n

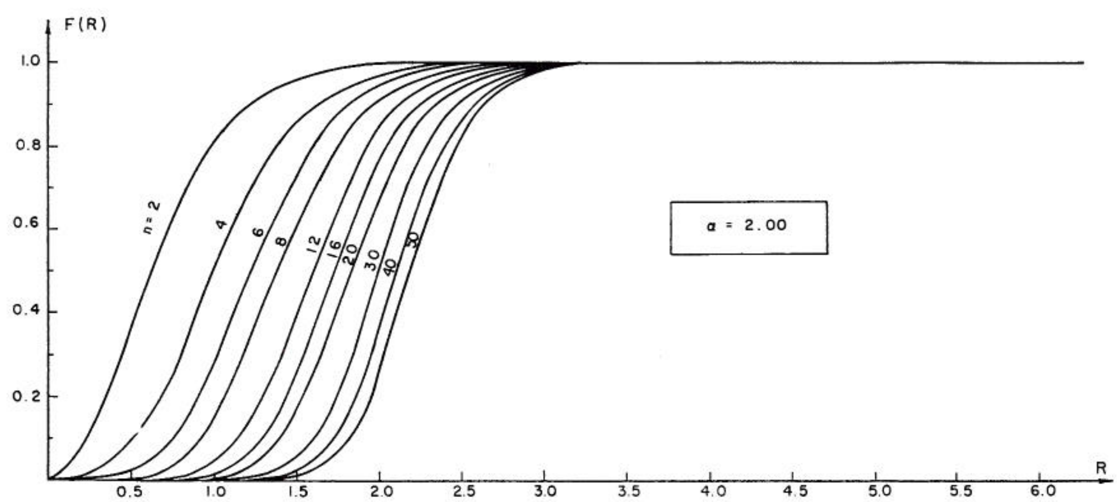

Fig. 3.9 Distribution functions of range, F (R), for α = 2.00 and various values of n

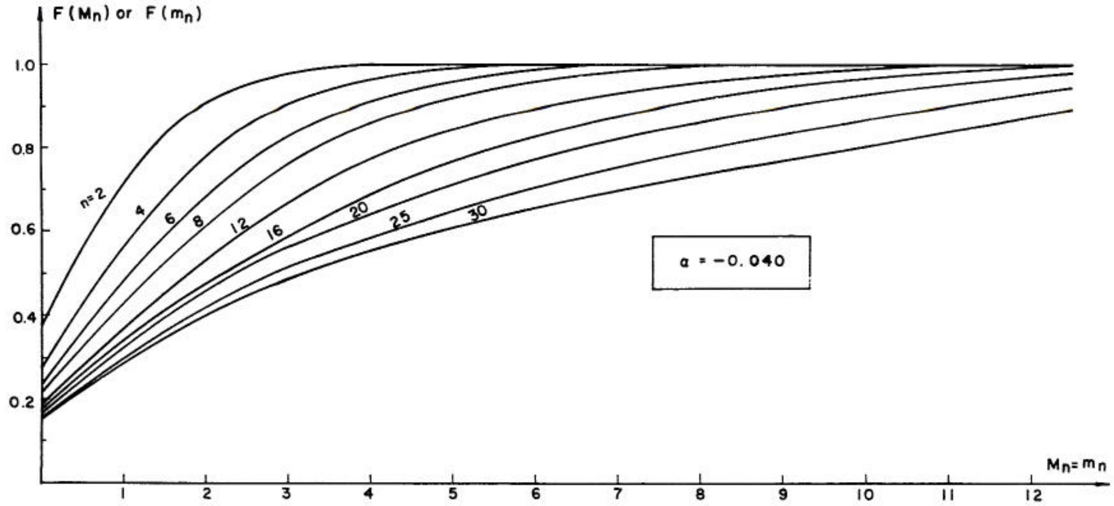

Fig. 3.10 Distribution functions of maximum surplus, $F(M_n)$, for α = -0.040 and various values of n

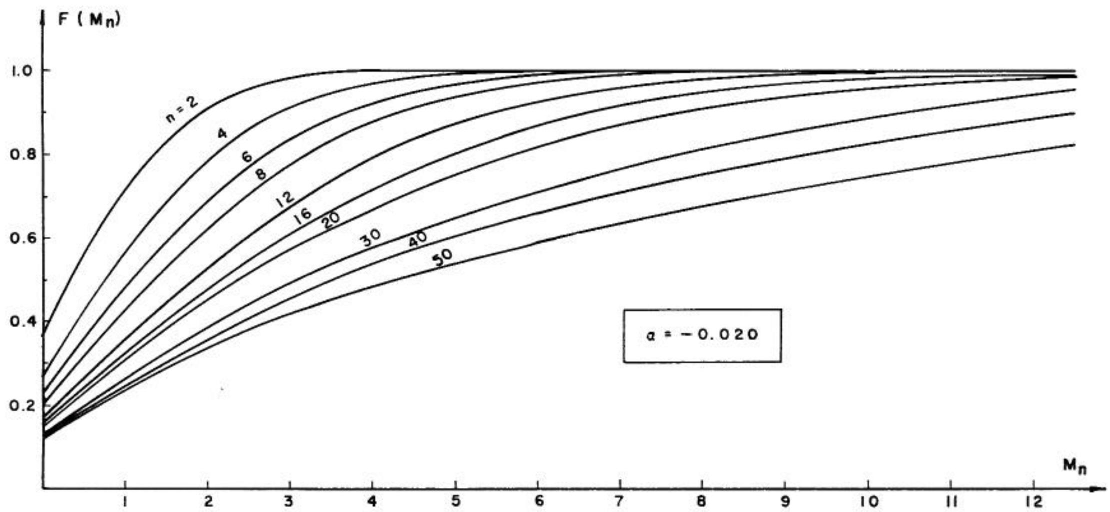

Fig. 3.11 Distribution functions of maximum surplus, $F(M_n)$, for α = -0.020 and various values of n

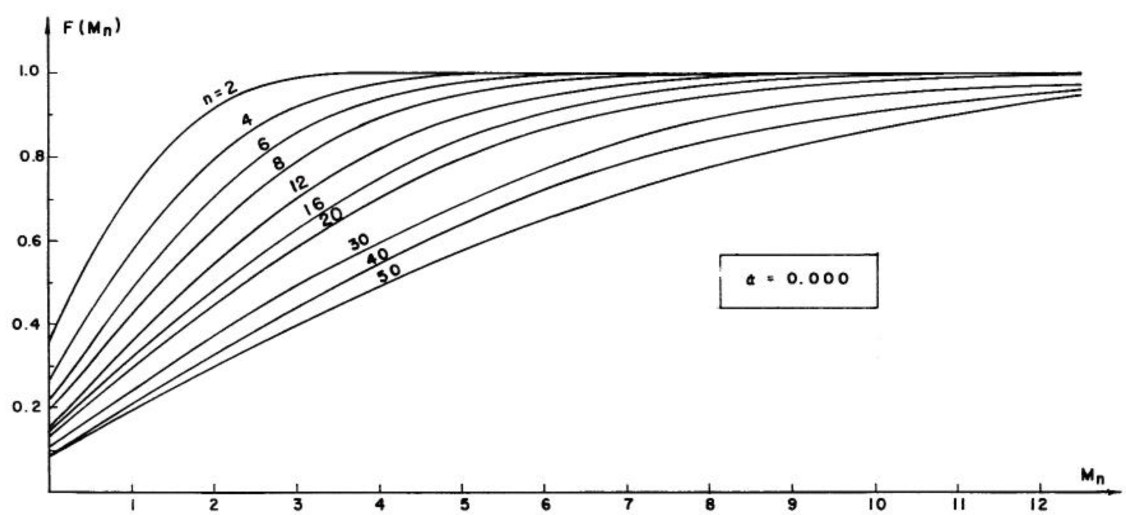

Fig. 3.12 Distribution functions of maximum surplus, $F(M_n)$, for α = 0.00 and various values of n

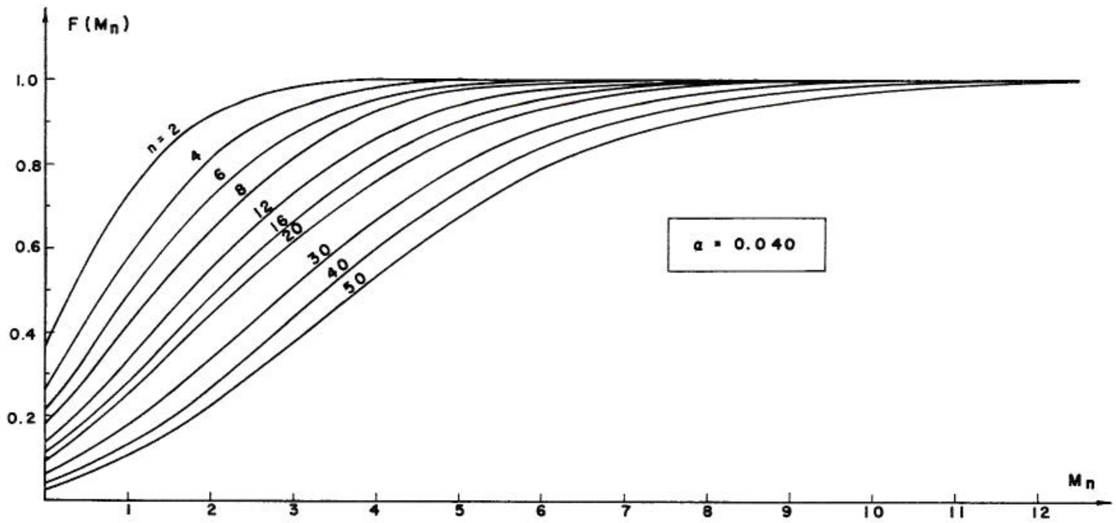

Fig. 3.13 Distribution functions of maximum surplus, $F(M_n)$, for $\alpha = 0.040$ and various values of n

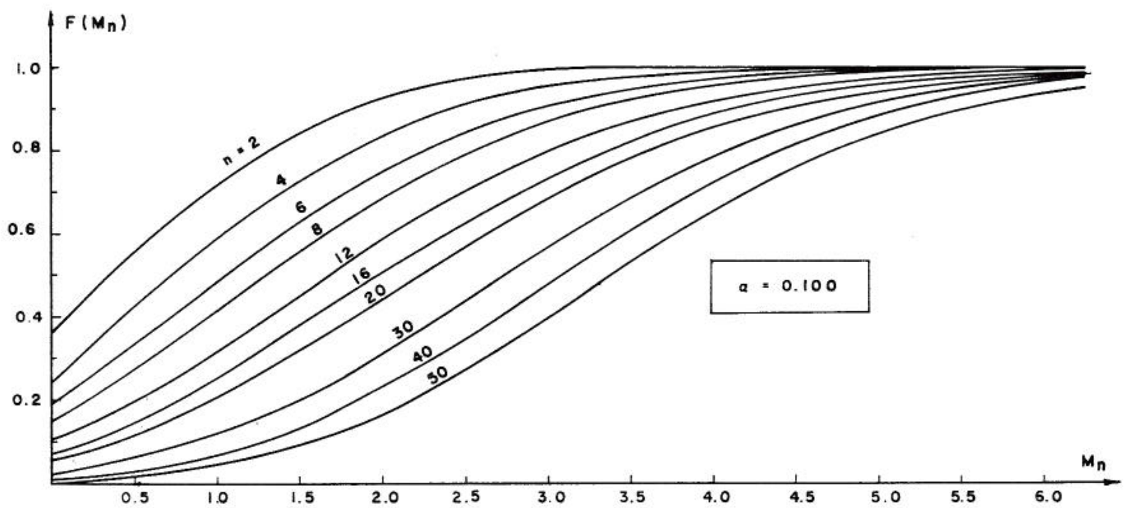

Fig. 3.14 Distribution functions of maximum surplus, $F(M_n)$, for $\alpha = 0.100$ and various values of n

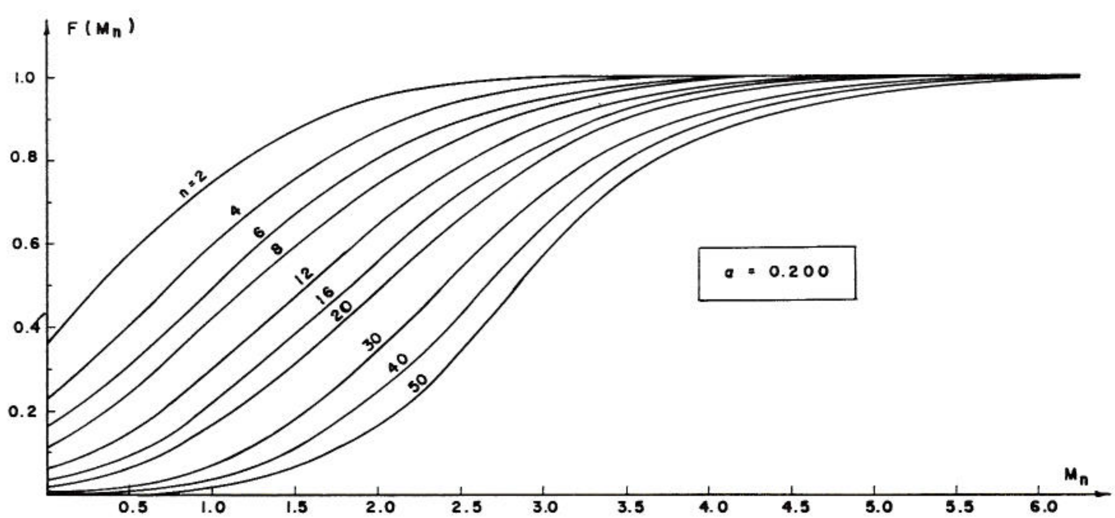

Fig. 3.15 Distribution functions of maximum surplus, $F(M_n)$, for $\alpha = 0.200$ and various values of n

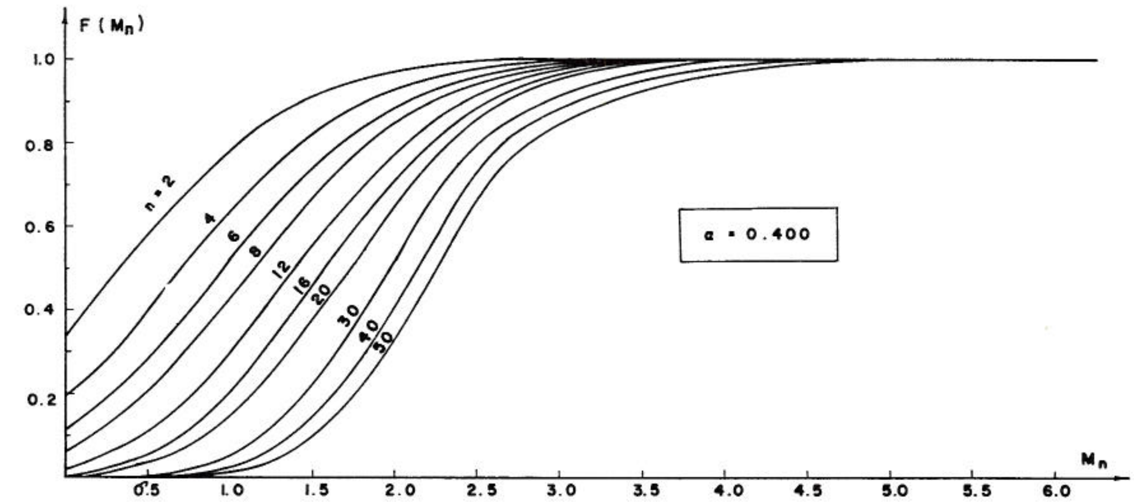

Fig. 3.16 Distribution functions of maximum surplus, $F(M_n)$, for α = 0.400 and various values of n

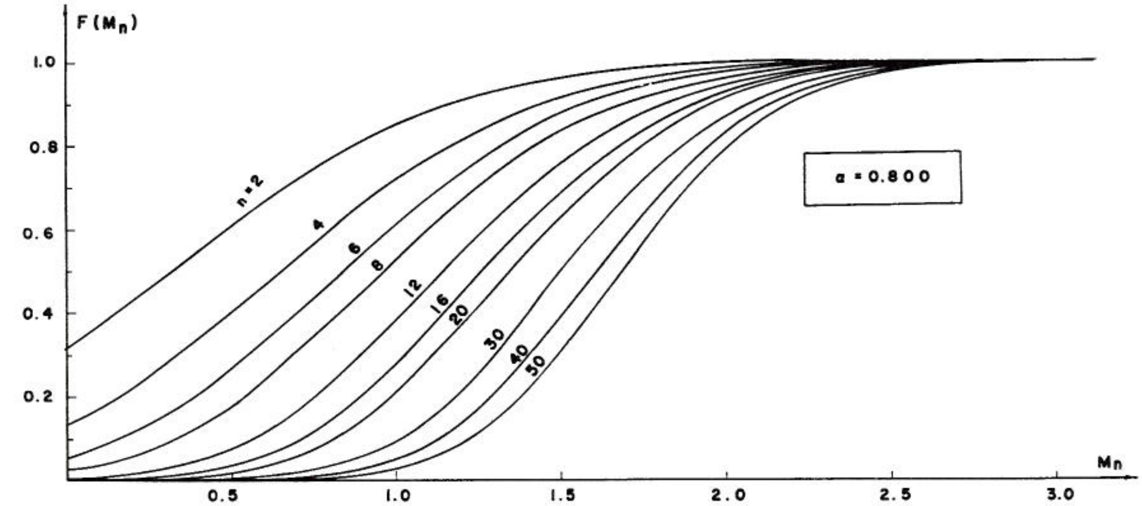

Fig. 3.17 Distribution functions of maximum surplus, $F(M_n)$, for α = 0.800 and various values of n

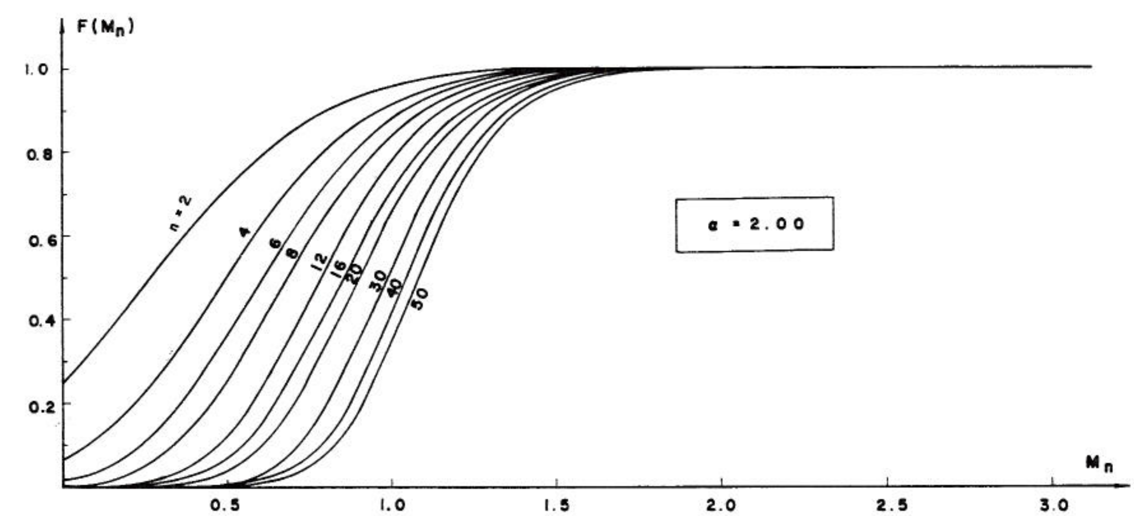

Fig. 3.18 Distribution functions of maximum surplus, $F(M_n)$, for α = 2.00 and various values of n

CHAPTER V

CONCLUSIONS

The following conclusions are based on the analysis of the range when the output is linearly dependent on the cumulative sum:

(1) For $\alpha \neq 0$, it is practically impossible to derive a general theoretical expression for the range distributions or their moments even for very small values of n, such as 2 or 3.

(2) The agreement between results obtained by the data generation method on a digital computer and the theory derived in this study, indicates that the hypotheses tested in the development of theory are confirmed.

(3) Probability density functions of the cumulative sum at the time t become approximately stationary for $t \geq 2.3/\alpha$.

(4) Equations 1.61 and 1.62 gives the mean range and the variance of range, respectively. They are derived under the hypothesis that the mean range and the variance of range depend only on the variance of the cumulative sum at the time t for a given α. As α tends to zero the mean range tends to the same expression as it is found by Anis and Lloyd and the variance tends to the same expression as derived by Feller.

(5) The probability density function for the range for $t > 5/\alpha$ is approximately normal, with the mean as shown in fig. 2.5 and the variance as shown in fig. 2.7.

(6) The general expression for the correlation coefficients between the maximum surplus (M_n) and the maximum deficit (m_n) of the cumulative sums is related to the variance of R_n, but the values of these correlation coefficients are obtained on a digital computer by the data generation method.

(7) The mean range and the variance of range decrease rapidly with an increase of α, while the variance of the output increases with an increase of α.

(8) The variance of the basic input variable is by far the most important parameter which affects the characteristics of range, surplus and deficit for given values of parameters α and n.

BIBLIOGRAPHY

1. Anis, A. A., and Lloyd, E. H., On the range of partial sums of a finite number of independent normal variates. Biometrika, Vol. 40; pp. 35-42, 1953.

2. Anis, A. A., The variance of the maximum of partial sums of a finite number of independent normal variates. Biometrika, Vol. 42; pp. 96-101, 1955.

3. Anis, A. A., On the moments of the maximum of partial sums of a finite number of independent normal variates. Biometrika, Vol. 43; pp. 79-84, 1956.

4. Feller, W., The asymptotic distributions of the range of sums of independent random variables. Annals of Mathematical Statistics, Vol. 22; pp. 427-432, 1951.

5. Hurst, H. E., Long term storage capacity of reservoirs. Transactions of the American Society of Civil Engineers, Vol. 116; pp. 770-808, 1951.

6. Melentijevich, M. J., Characteristics of storage when outflow is dependent upon reservoir volume. Ph. D. Dissertation, Colorado State University, 1965, 96 p.

7. Thomas, A. H., Operation research in water quality management. Harvard University Division of Engineering and Applied Physics, Chapter I; pp. 1-26, 1963.

8. Wax, N., Selected papers on noise and stochastic processes. New York, Dover Publications, 1954, 337 p.

Key Words: Reservoir design, Water storage problems, Range of the cumulative sums S_t.

Abstract: The objectives of this investigation are twofold: (i) Determination of the distribution function for the cumulative sums S_t when the output is dependent upon these sums; and (ii) Development of probability expressions for the range of cumulative departures of a stochastic variable. The basic relationship between input, output and cumulative sum is expressed by the equation of continuity $dS/dt = q_i(t) - \alpha S$, with α = constant. The following principal assumptions concern the fluctuating portion $q_i(t)$: (i) $q_i(t)$ is a normal independent variable with a mean of zero and a variance σ^2; (ii) The correlation between the values of $q_i(t)$ at different times t_1 and t_2 exists only when $|t_1 - t_2|$ is very small; and (iii) $q_i(t)$ varies an extreme rapid amount when compared with the variation of the cumulative sum S_t. The distribution function for the cumulative sums S_t when the output is dependent upon these sums is defined by

$$f(S, t; S_o) = \frac{\sqrt{\alpha}}{[\pi(1-e^{-2\alpha t})\sigma^2]^{1/2}} e^{-\alpha(S - S_o e^{-2\alpha t})^2/(1-e^{-2\alpha t})\sigma^2} .$$

Key Words: Reservoir design, Water storage problems, Range of the cumulative sums S_t.

Abstract: The objectives of this investigation are twofold: (i) Determination of the distribution function for the cumulative sums S_t when the output is dependent upon these sums; and (ii) Development of probability expressions for the range of cumulative departures of a stochastic variable. The basic relationship between input, output and cumulative sum is expressed by the equation of continuity $dS/dt = q_i(t) - \alpha S$, with α = constant. The following principal assumptions concern the fluctuating portion $q_i(t)$: (i) $q_i(t)$ is a normal independent variable with a mean of zero and a variance σ^2; (ii) The correlation between the values of $q_i(t)$ at different times t_1 and t_2 exists only when $|t_1 - t_2|$ is very small; and (iii) $q_i(t)$ varies an extreme rapid amount when compared with the variation of the cumulative sum S_t. The distribution function for the cumulative sums S_t when the output is dependent upon these sums is defined by

$$f(S, t; S_o) = \frac{\sqrt{\alpha}}{[\pi(1-e^{-2\alpha t})\sigma^2]^{1/2}} e^{-\alpha(S - S_o e^{-2\alpha t})^2/(1-e^{-2\alpha t})\sigma^2} .$$

Key Words: Reservoir design, Water storage problems, Range of the cumulative sums S_t.

Abstract: The objectives of this investigation are twofold: (i) Determination of the distribution function for the cumulative sums S_t when the output is dependent upon these sums; and (ii) Development of probability expressions for the range of cumulative departures of a stochastic variable. The basic relationship between input, output and cumulative sum is expressed by the equation of continuity $dS/dt = q_i(t) - \alpha S$, with α = constant. The following principal assumptions concern the fluctuating portion $q_i(t)$: (i) $q_i(t)$ is a normal independent variable with a mean of zero and a variance σ^2; (ii) The correlation between the values of $q_i(t)$ at different times t_1 and t_2 exists only when $|t_1 - t_2|$ is very small; and (iii) $q_i(t)$ varies an extreme rapid amount when compared with the variation of the cumulative sum S_t. The distribution function for the cumulative sums S_t when the output is dependent upon these sums is defined by

$$f(S, t; S_o) = \frac{\sqrt{\alpha}}{[\pi(1-e^{-2\alpha t})\sigma^2]^{1/2}} e^{-\alpha(S - S_o e^{-2\alpha t})^2/(1-e^{-2\alpha t})\sigma^2} .$$

Key Words: Reservoir design, Water storage problems, Range of the cumulative sums S_t.

Abstract: The objectives of this investigation are twofold: (i) Determination of the distribution function for the cumulative sums S_t when the output is dependent upon these sums; and (ii) Development of probability expressions for the range of cumulative departures of a stochastic variable. The basic relationship between input, output and cumulative sum is expressed by the equation of continuity $dS/dt = q_i(t) - \alpha S$, with α = constant. The following principal assumptions concern the fluctuating portion $q_i(t)$: (i) $q_i(t)$ is a normal independent variable with a mean of zero and a variance σ^2; (ii) The correlation between the values of $q_i(t)$ at different times t_1 and t_2 exists only when $|t_1 - t_2|$ is very small; and (iii) $q_i(t)$ varies an extreme rapid amount when compared with the variation of the cumulative sum S_t. The distribution function for the cumulative sums S_t when the output is dependent upon these sums is defined by

$$f(S, t; S_o) = \frac{\sqrt{\alpha}}{[\pi(1-e^{-2\alpha t})\sigma^2]^{1/2}} e^{-\alpha(S - S_o e^{-2\alpha t})^2/(1-e^{-2\alpha t})\sigma^2} .$$