CHARTS OF SOME UPPER PERCENTAGE POINTS OF THE DISTRIBUTION OF THE LARGEST CHARACTERISTIC ROOT[1]

BY D. L. HECK[2]

University of North Carolina

1. Introduction. In multivariate analysis, the largest characteristic root of certain matrices of sample quantities, or a simple function of this root, provides a statistic for testing (i) independence between a set of p correlated variates and a set of q correlated variates in a $(p + q)$-variate normal population and (ii) the general multivariate linear hypothesis, assuming multivariate normal populations. Likelihood ratio methods for dealing with these tests have also been advanced by Wilks [29] and Bartlett [4], and comprehensive accounts of the use of these techniques are given by Wilks [30], Rao [17], and Anderson [2].

Procedures based on the largest characteristic root were proposed by Roy [19, 20, 23] for not only testing (i) and (ii), but also for obtaining confidence bounds on parametric functions associated with both cases. These procedures require the c.d.f. of the largest root, which was given in terms of a chain of recursion formulae by Roy [19] and Nanda [12] who started from the joint sampling distribution of the roots obtained earlier by Fisher [6], Girshick [10], Hsu [11], and Roy [18]. This distribution of θ_i $(i = 1, 2, \cdots, s)$, the s non-zero roots obtained under the null hypothesis in (i) and (ii), is given by

$$p(\theta_1, \theta_2, \cdots, \theta_s) \prod_{i=1}^{s} d\theta_i$$

$$= \frac{\pi^{s/2} \prod_{i=1}^{s} \Gamma\left(\frac{2m + 2n + s + i + 2}{2}\right) \prod_{i=1}^{s} \theta_i^m (1 - \theta_i)^n \prod_{i>j}^{s} (\theta_i - \theta_j) \prod_{i=1}^{s} d\theta_i}{\prod_{i=1}^{s} \Gamma\left(\frac{2m + i + 1}{2}\right) \Gamma\left(\frac{2n + i + 1}{2}\right) \Gamma\left(\frac{i}{2}\right)}$$

$$0 < \theta_1 \leqq \cdots \leqq \theta_s < 1,$$

where θ_i and the parameters s, m, and n assume different values, depending upon the hypothesis being tested.

For example, in (i) the θ_i's are the s non-zero roots of the $(p \times p)$ matrix $S_{12}S_{22}^{-1}S_{12}^{1}S_{11}^{-1}$, where

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{12}' & S_{22} \end{bmatrix} \begin{matrix} p \\ q \end{matrix}$$
$$\begin{matrix} p & q \end{matrix}$$

Received April 29, 1958; revised May 25, 1959.

[1] This paper was prepared with the support of the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command. Reproduction in whole or part is permitted for any purpose of the United States Government.

[2] Now with International Business Machines Corporation, Endicott, New York.

is the sample covariance matrix based on $N - 1$ degrees of freedom. The parameters are given by $s = \min(p, q)$, $m = (|p - q| - 1)/2$, and

$$n = (N - p - q - 2)/2.$$

In (ii) and the special case of testing for equality of the $(p \times 1)$ mean vectors of k groups in one-way classification multivariate analysis of variance, $\theta_i = c_i/(1 + c_i)$, where the c_i's are the s non-zero roots of the $(p \times p)$ matrix BW^{-1}. The elements of the $(p \times p)$ matrix B consist of the "between groups" corrected sums of squares and crossproducts of the p variates, while W is the corresponding matrix for "within groups". With n_j observations on the p variates in each of the k groups, $s = \min(k - 1, p)$, $m = (|k - p - 1| - 1)/2$, and $n = (\sum_{j=1}^{k} n_j - k - p - 1)/2$.

The tests are carried out by computing θ_s and then comparing this statistic with the appropriate 100α percentage point of the distribution of the largest root. Further applications of these percentage points for testing purposes have been given by Bargmann [3], Chaudhuri [5], Foster and Rees [7], Pillai [15], Roy [20], [23], Roy and Bargmann [25], and Roy and Roy [27]. In addition, the use of the points in setting up multivariate confidence bounds is discussed in [3], [21], [22], [23], [24], [25], [26], [27].

The purpose of this paper is to present, in chart form, some upper percentage points of the distribution of the largest root for a wider range of the parameters than has heretofore been considered. One of the most extensive tabulations to date is that by Pillai [15], giving the upper 1% and 5% points and covering the range $s = 2(1)5$, $m = 0(1)4$, $n = 5(5)40(20)100(30)160, 200, 300, 500, 1000$. Also, the upper 1% and 5% points for these same values of m and n have been obtained by Pillai and Bantegui [16] for $s = 6$. Other tables include Nanda's [13], the upper 1% and 5% points for $s = 2$, $m = 0(\frac{1}{2})2$, $n = \frac{1}{2}(\frac{1}{2})10$; Chaudhuri's [5], the upper 1% and 5% points for $s = 2$, $m = n = 2\frac{1}{2}(\frac{1}{2})5(1)11$, for $s = 3$, $m = n = 2\frac{1}{2}(\frac{1}{2})5(1)8$, and for $s = 3$, $m = 0(\frac{1}{2})2$, $n = \frac{1}{2}(\frac{1}{2})2$; Foster and Rees' [7], the upper 1%, 5%, 10%, 15%, and 20% points for $s = 2$, $m = -\frac{1}{2}$, $0(1)9$, $n = 1(1)19(5)49, 59, 79$; and Foster's [8, 9], the upper 1%, 5%, 10%, 15%, and 20% points for $s = 3, 4$, $m = -\frac{1}{2}(\frac{1}{2})3$, $n = 0(1)95$. From Table 4.1 and the charts in Section 3, the upper 1%, 2.5%, and 5% points may be obtained for $s = 2(1)5$, $m = -\frac{1}{2}, 0(1)10$, $n \geqq 5$.

2. Computation of the percentage points. The charts in Section 3 were prepared from percentage points which were computed using two types of approximations to the c.d.f. of the largest characteristic root. The first type of approximation, obtained for $s = 2, 3, 4, 5$ by Pillai [14] was used to compute, in general, the points for integral $n \leqq 100$. For large values of n, generally $n > 100$ asymptotic approximations based on Pillai's formulae were used which were obtained by Whittlesey [28].

To compute the percentage points from Pillai's approximations, denoted by $p_s(x, m, n)$, the value of $p_s(x, m, n)$ for a particular combination (s, m, n) was first calculated at the 100 values of x from .01 to 1.0 at intervals of .01. On the resulting ordinates, a method of inverse interpolation was used to obtain the upper 1%, 2.5%, and 5% points, i.e. x_α such that

$$p_s(x_\alpha, m, n) = 1 - \alpha \qquad (\alpha = .01, .025, .05).$$

The overall computational procedure for each value of s was as follows: For a fixed integral m and an initial (small) n, the percentage points were computed; n was then stepped up by unit increments, with the percentage points being computed for each value of n, until the desired set of values of n was covered. Then the expression was modified for the next integral value of m, and the percentage points for this value of m were computed for all desired n. This procedure was continued to $m = 10$, which is a fairly large value for practical purposes.

As a check on the accuracy of these percentage points, a number of the points were substituted in the expression for the exact c.d.f., and the largest error which occurred was found to be less than two units in the fourth decimal.

Whittlesey's asymptotic approximations (for integral values of [illegible]) obtained from Pillai's approximations by using Stirling's approximat[illegible] and the substitution

$$z = -(m + 2n + s + 1) \log (1 - x), \tag{2.1}$$

and then letting n become large. From the resulting expressions, denoted by $w_s(z, m)$, inverse interpolation was used to obtain $z_\alpha(s, m)$ (or z_α) such that for fixed s and m,

$$w_s(z_\alpha, m) = 1 - \alpha \qquad (\alpha = .01, .025, .05).$$

From these "asymptotic" $z_\alpha(s, m)$ values, given in Table 4.1, the percentage points $x_\alpha(s, m, n)$ were obtained by inverting (2.1).

A group of the percentage points obtained from Whittlesey's approximations was checked by substitution in the expression for the exact c.d.f., and of those points used in the final tabulation, the error for the most unfavorable combination of s and m ($s = 5, m = 10$) was found to be five units in the fourth decimal. This error, which is primarily an error of asymptotic approximation, is considerably smaller for smaller values of s and m, and, because of the asymptotic nature of the approximation, decreases in all cases, for increasing n.

Computation of the percentage points and the $z_\alpha(s, m)$ values was carried out on the IBM 650, with the programs coded in The Bell Interpretive System [31]. The program of the exact c.d.f. of the largest root, which was used for checking purposes, was coded in DOPSIR [1] (for $s = 2(1)6$, $m = 0(1)10$, and integral $n \geqq 0$), and is available at the North Carolina State College IBM Laboratory, The Institute of Statistics, Raleigh, North Carolina. The computation of the points for $m = -\frac{1}{2}$ was done subsequent to the computation for integral valued m, and Pillai's and Whittlesey's approximations were again used, after appropriate modifications were made.

CHART I

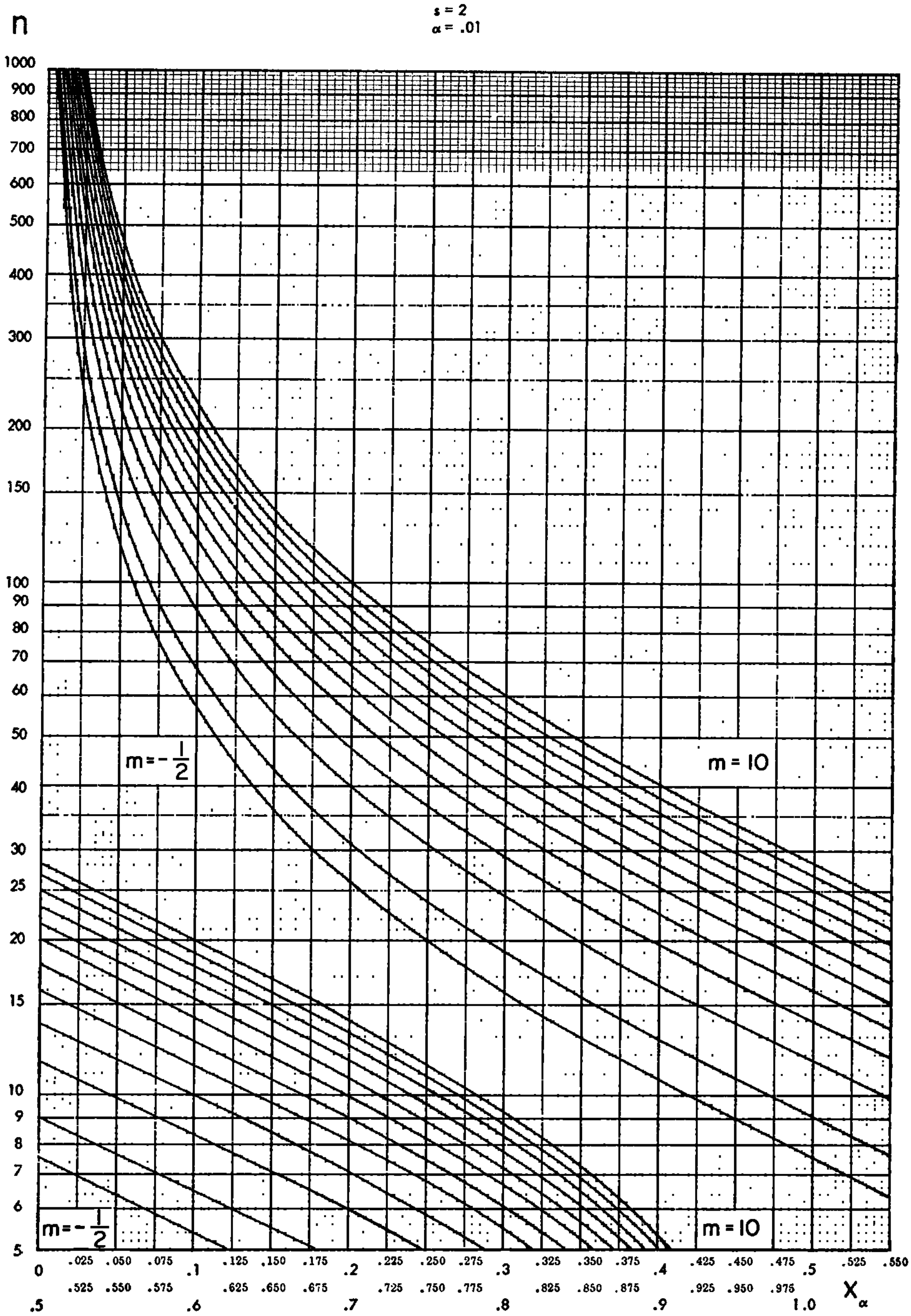

n
s = 2
α = .01
1000
900
800
700
600
500
400
300
200
150
100
90
80
70
60
50
40
30
25
20
15
10
9
8
7
6
5
m = -1/2
m = 10
m = -1/2
m = 10
0 .025 .050 .075 .1 .125 .150 .175 .2 .225 .250 .275 .3 .325 .350 .375 .4 .425 .450 .475 .5 .525 .550
.525 .550 .575 .625 .650 .675 .725 .750 .775 .825 .850 .875 .925 .950 .975
.5 .6 .7 .8 .9 1.0
X_α

CHART II

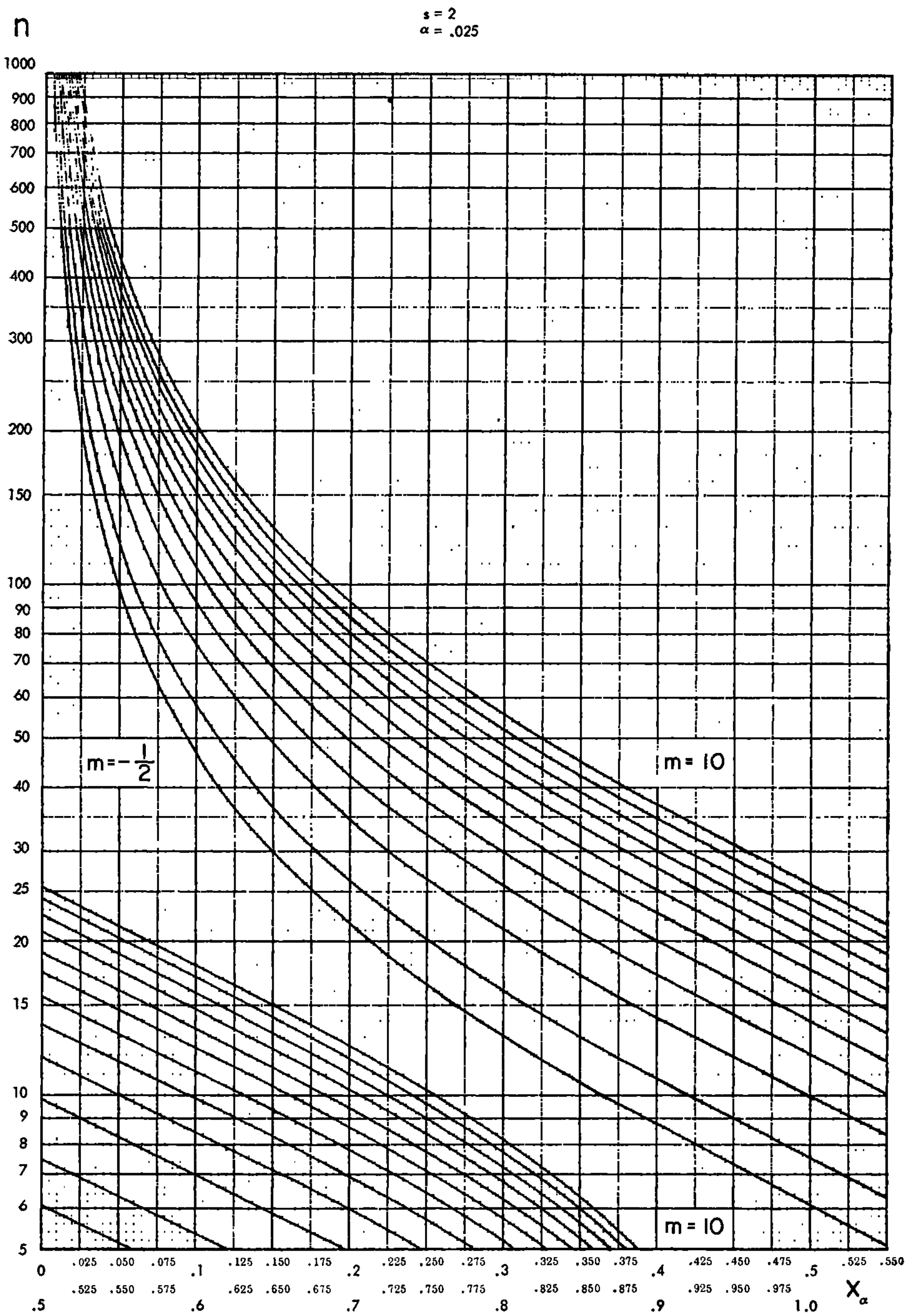

s = 2
α = .025
n
1000
900
800
700
600
500
400
300
200
150
100
90
80
70
60
50
40
30
25
20
15
10
9
8
7
6
5
m = −1/2
m = 10
m = 10
0 .025 .050 .075 .1 .125 .150 .175 .2 .225 .250 .275 .3 .325 .350 .375 .4 .425 .450 .475 .5 .525 .550
.5 .525 .550 .575 .6 .625 .650 .675 .7 .725 .750 .775 .8 .825 .850 .875 .9 .925 .950 .975 1.0
Xα

CHART III

s = 2
$\alpha = .05$

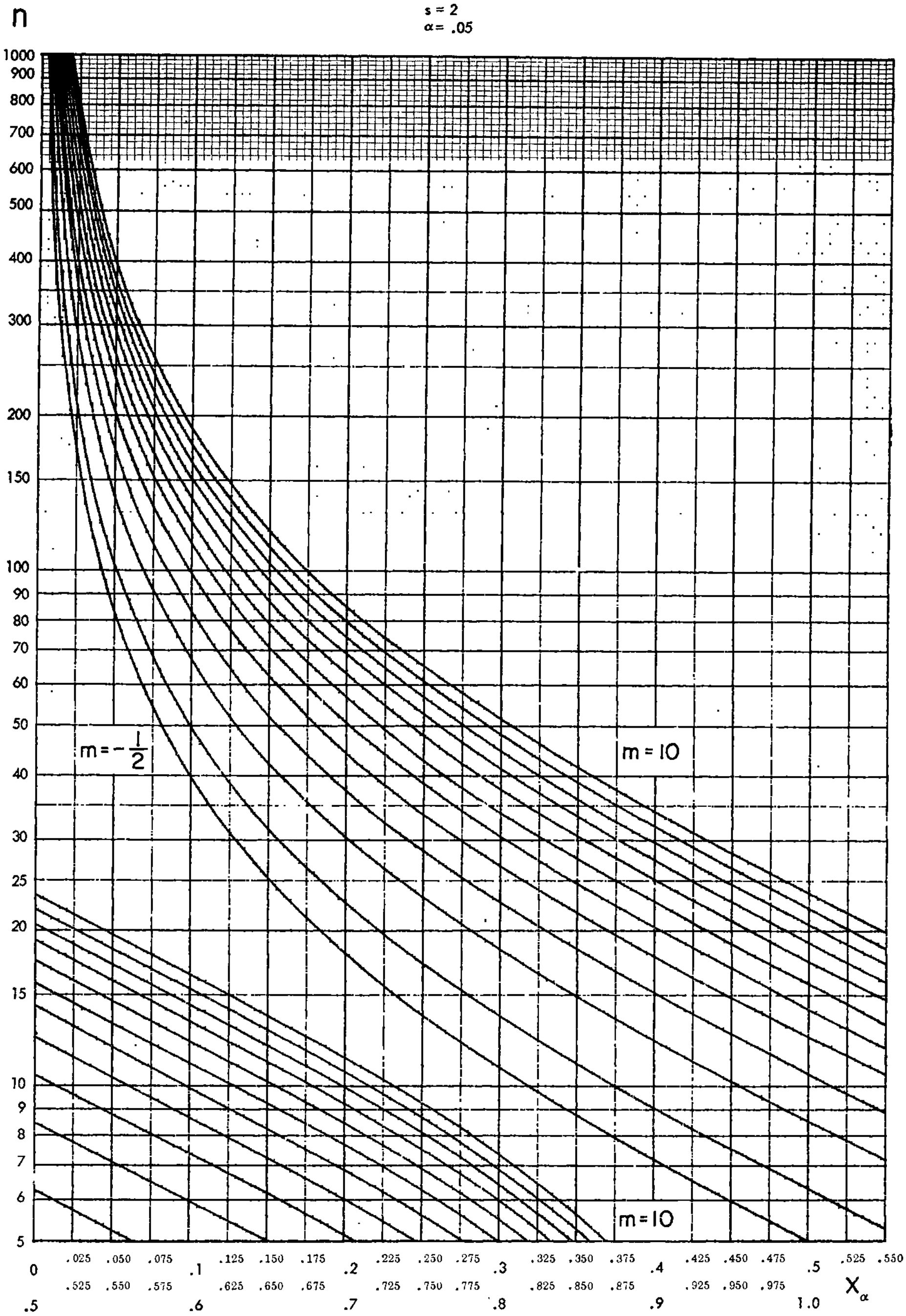


CHART IV

s = 3
α = .01

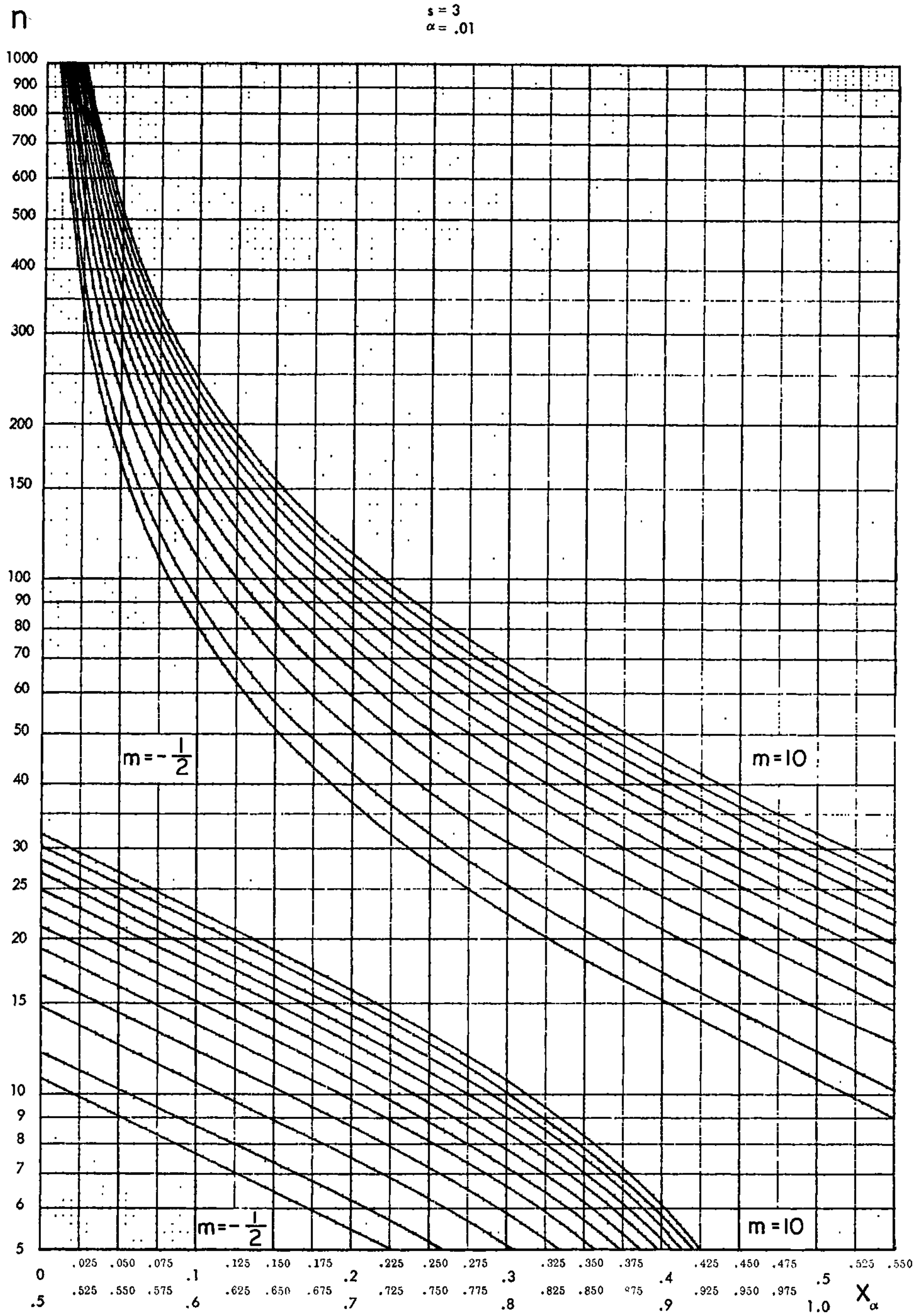


CHART V

$s = 3$
$\alpha = .025$

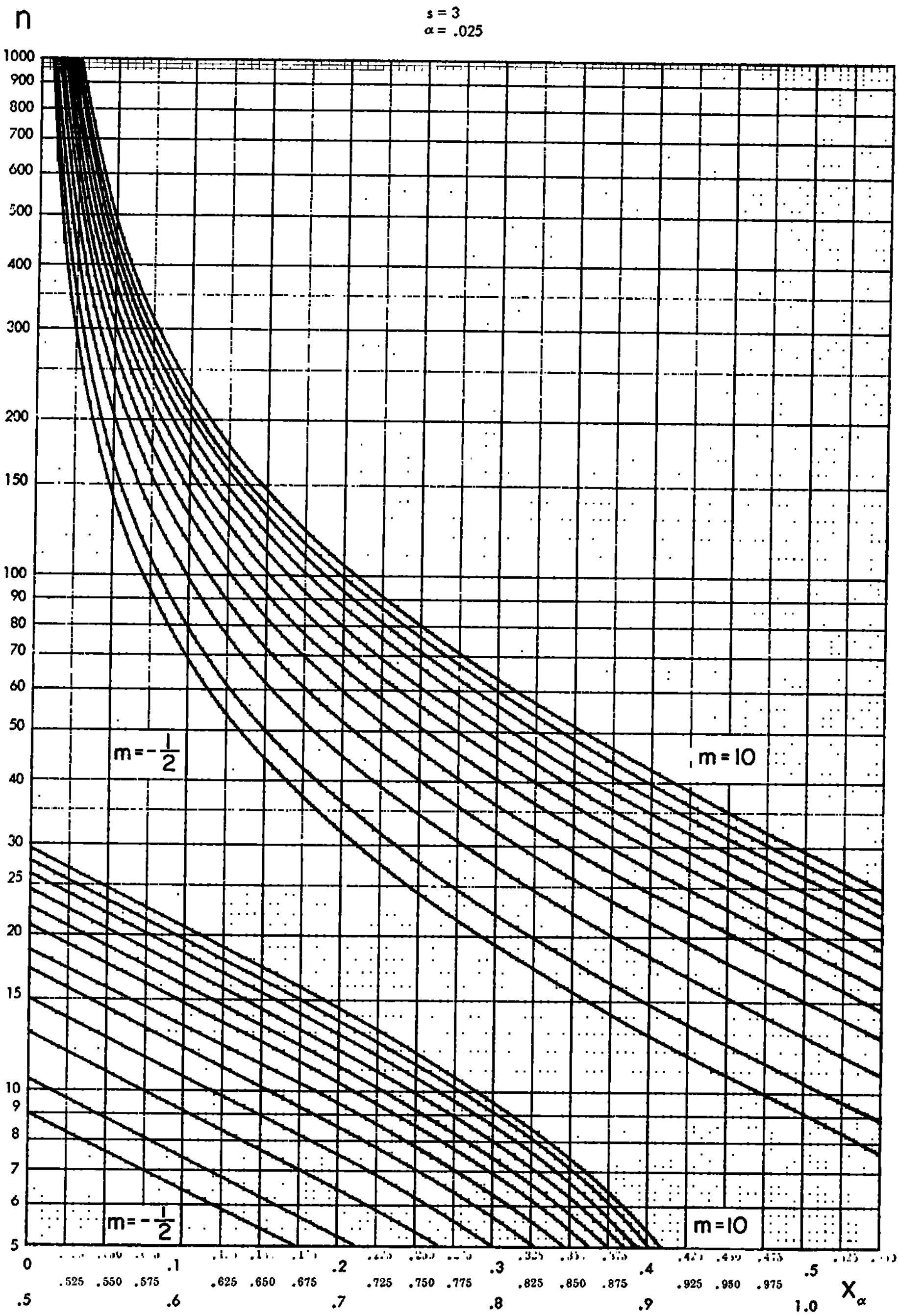


CHART VI

$s = 3$
$\alpha = .05$

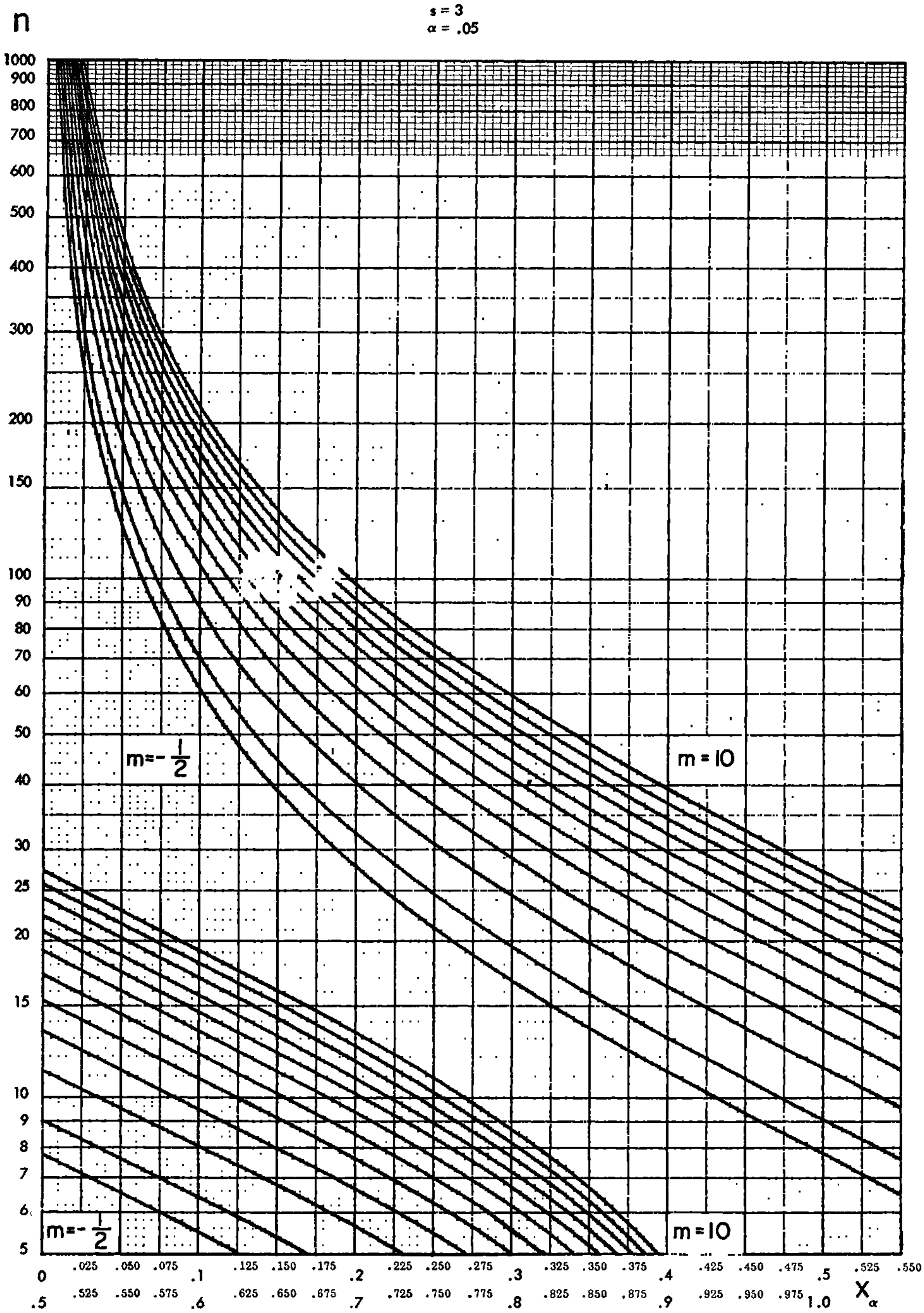


CHART VII

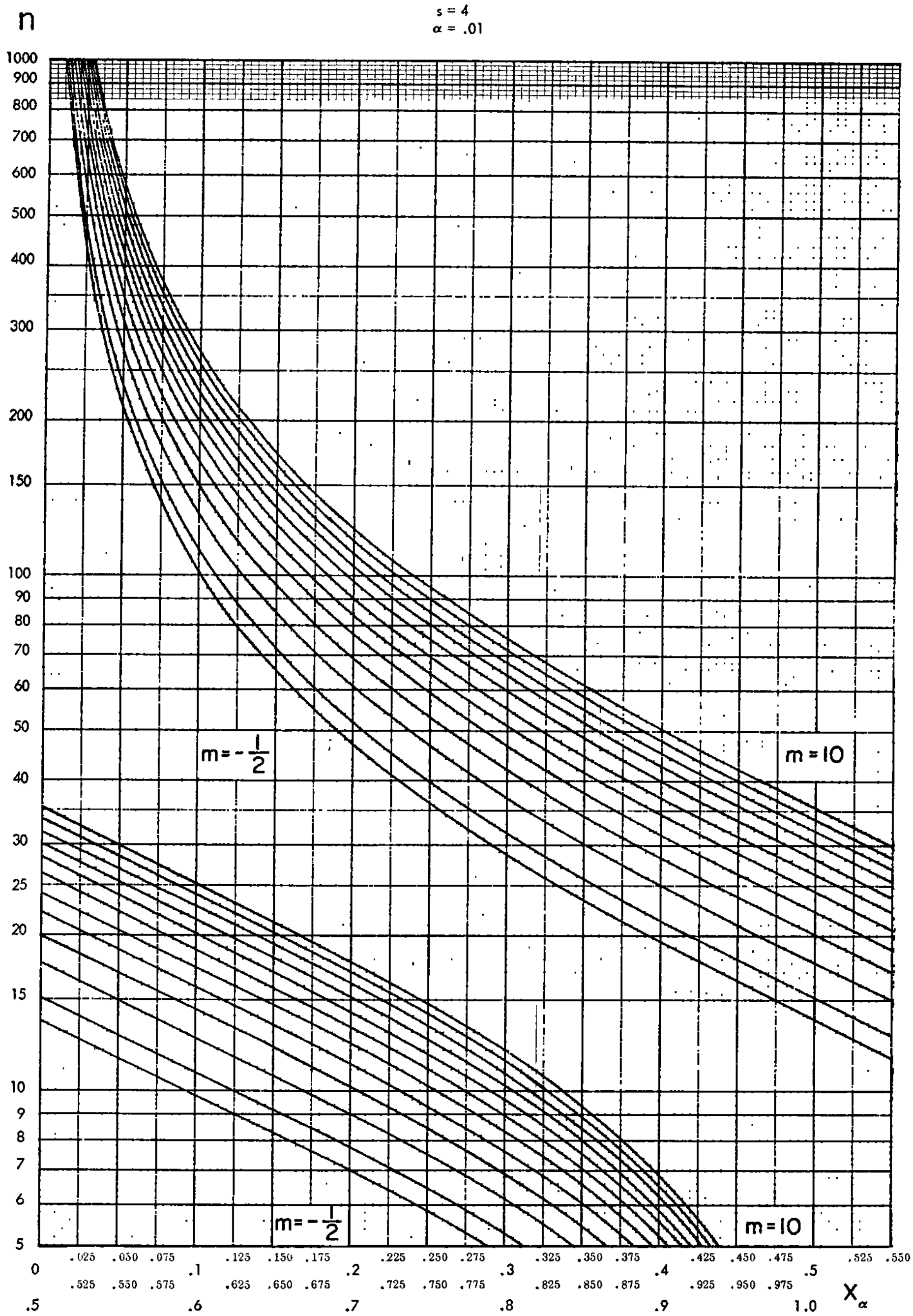

s = 4
α = .01
n
1000
900
800
700
600
500
400
300
200
150
100
90
80
70
60
50
40
30
25
20
15
10
9
8
7
6
5
m = -1/2
m = 10
m = -1/2
m = 10
0 .025 .050 .075 .1 .125 .150 .175 .2 .225 .250 .275 .3 .325 .350 .375 .4 .425 .450 .475 .5 .525 .550
.525 .550 .575 .625 .650 .675 .725 .750 .775 .825 .850 .875 .925 .950 .975
.5 .6 .7 .8 .9 1.0
X_α

CHART VIII

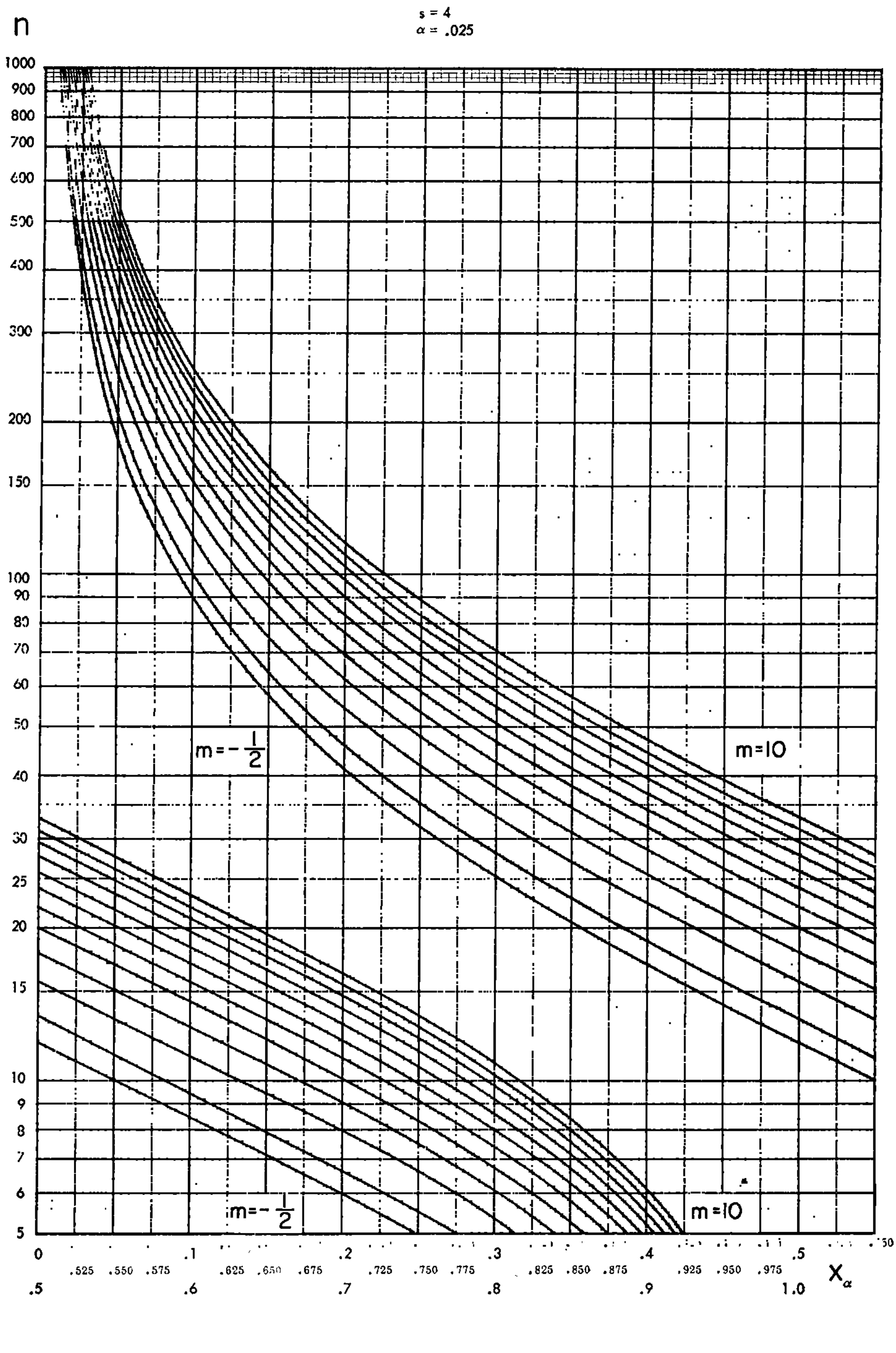

n
s = 4
α = .025
1000
900
800
700
600
500
400
300
200
150
100
90
80
70
60
50
40
30
25
20
15
10
9
8
7
6
5
m = -1/2
m=10
m = -1/2
m=10
0
.1
.2
.3
.4
.5
.50
.525
.550
.575
.625
.650
.675
.725
.750
.775
.825
.850
.875
.925
.950
.975
.5
.6
.7
.8
.9
1.0
Xα

CHART IX

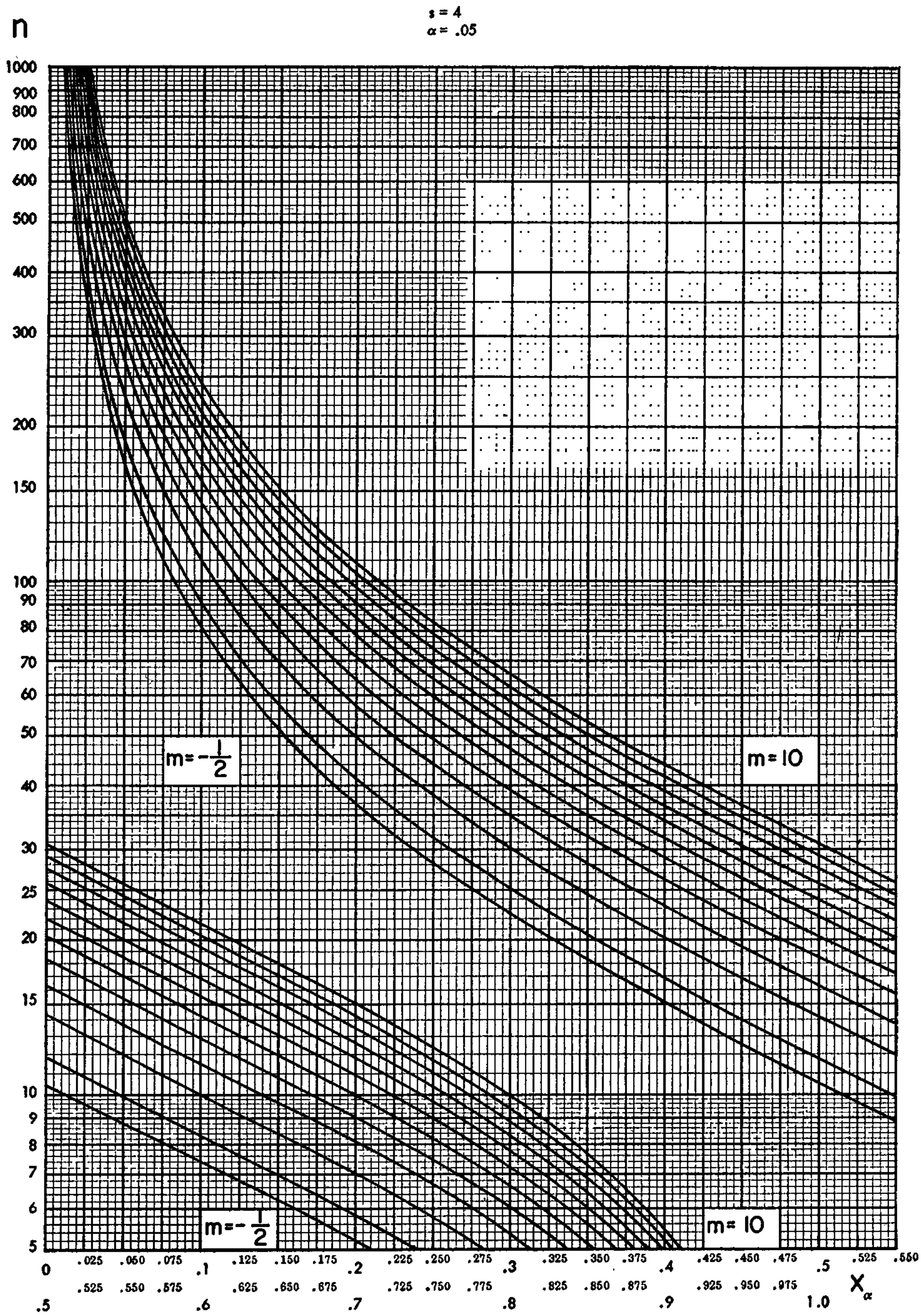

s = 4
α = .05
n
1000
900
800
700
600
500
400
300
200
150
100
90
80
70
60
50
40
30
25
20
15
10
9
8
7
6
5
m = −1/2
m = 10
m = −1/2
m = 10
0 .025 .050 .075 .1 .125 .150 .175 .2 .225 .250 .275 .3 .325 .350 .375 .4 .425 .450 .475 .5 .525 .550
.5 .525 .550 .575 .6 .625 .650 .675 .7 .725 .750 .775 .8 .825 .850 .875 .9 .925 .950 .975 1.0
X_α

CHART X

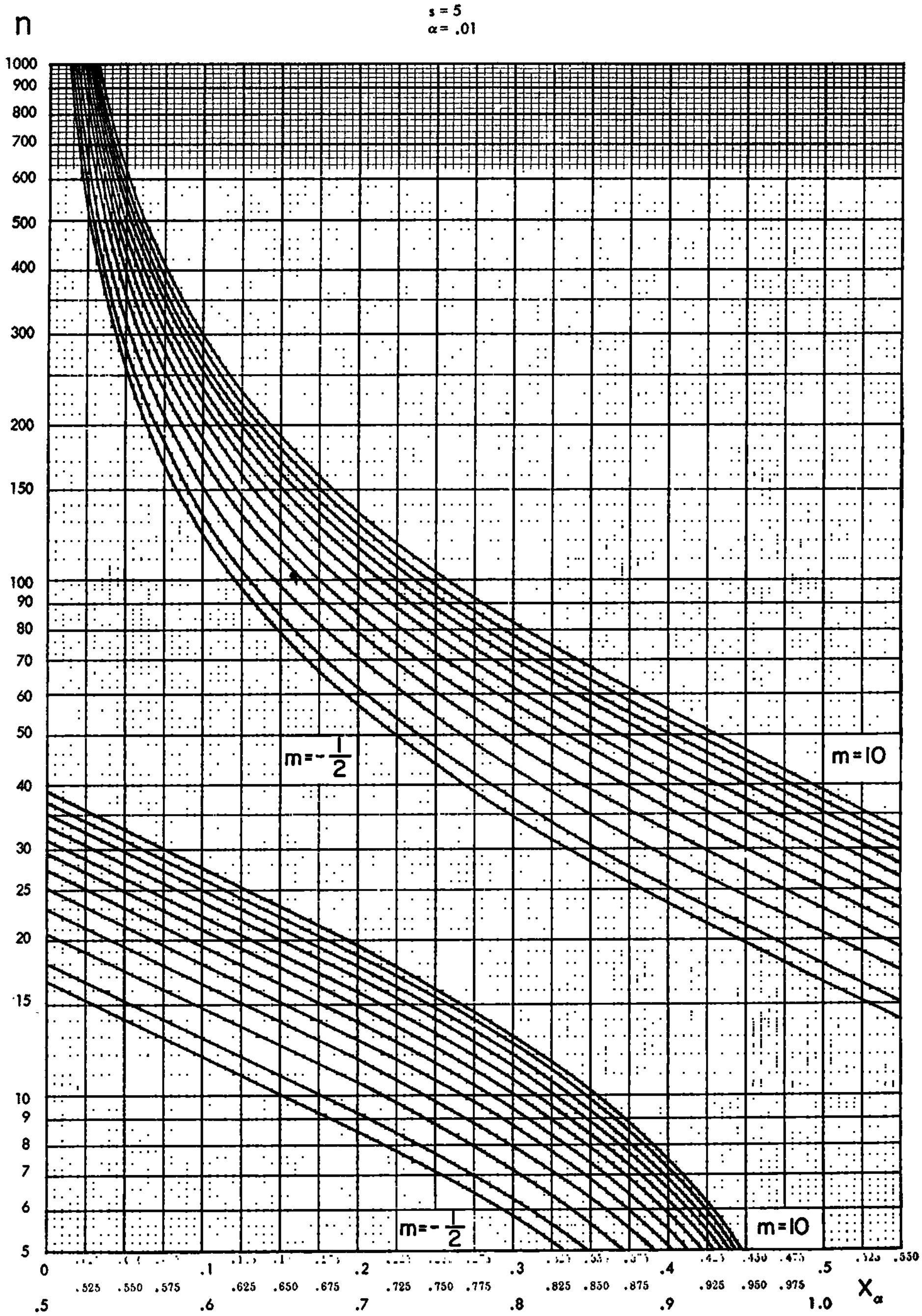

s = 5
α = .01
n
1000
900
800
700
600
500
400
300
200
150
100
90
80
70
60
50
40
30
25
20
15
10
9
8
7
6
5
m = −1/2
m = 10
m = −1/2
m = 10
0
.1
.2
.3
.4
.5
.550
.525
.550
.575
.625
.650
.675
.725
.750
.775
.825
.850
.875
.925
.950
.975
.5
.6
.7
.8
.9
1.0
X_α

CHART XI

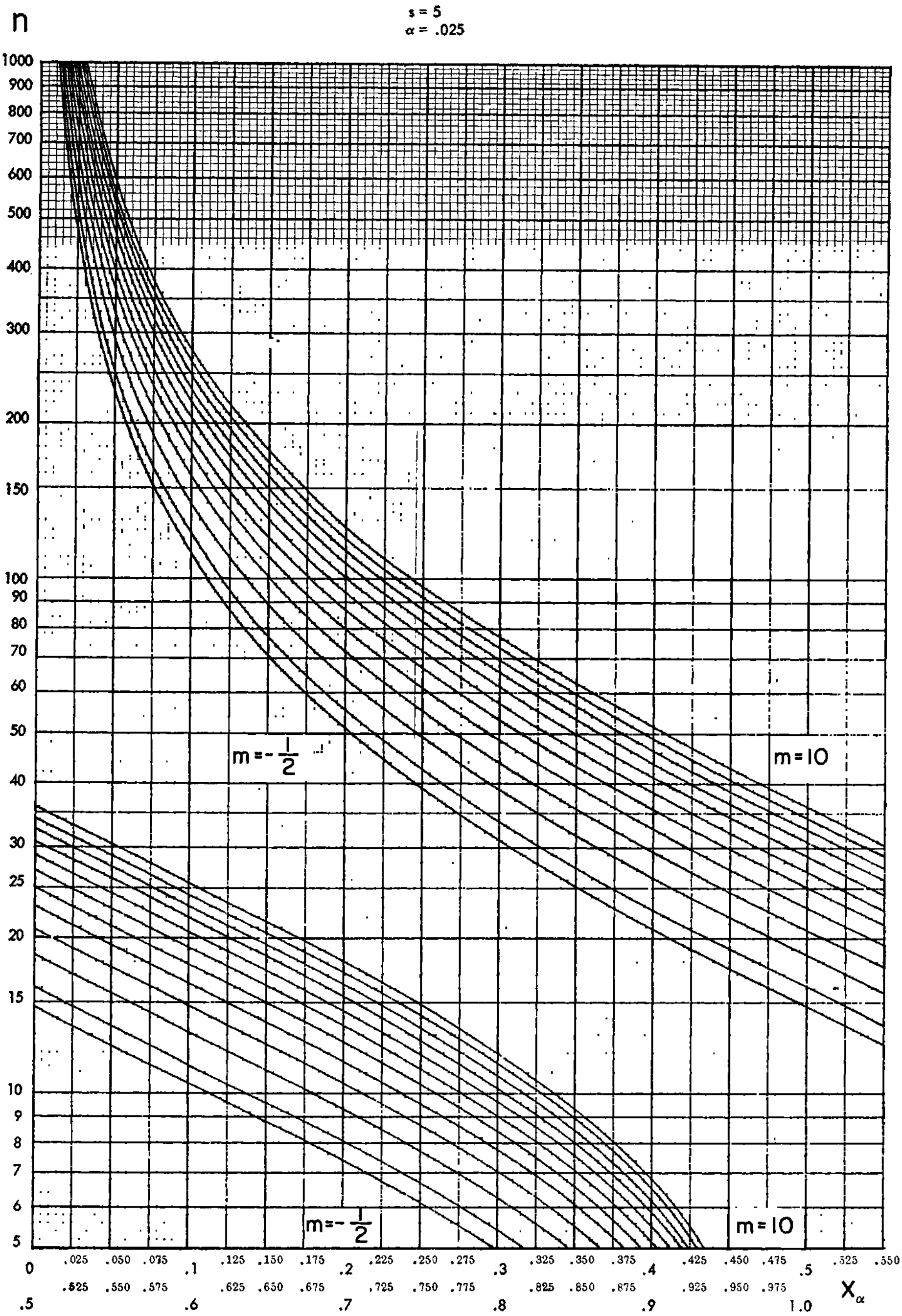
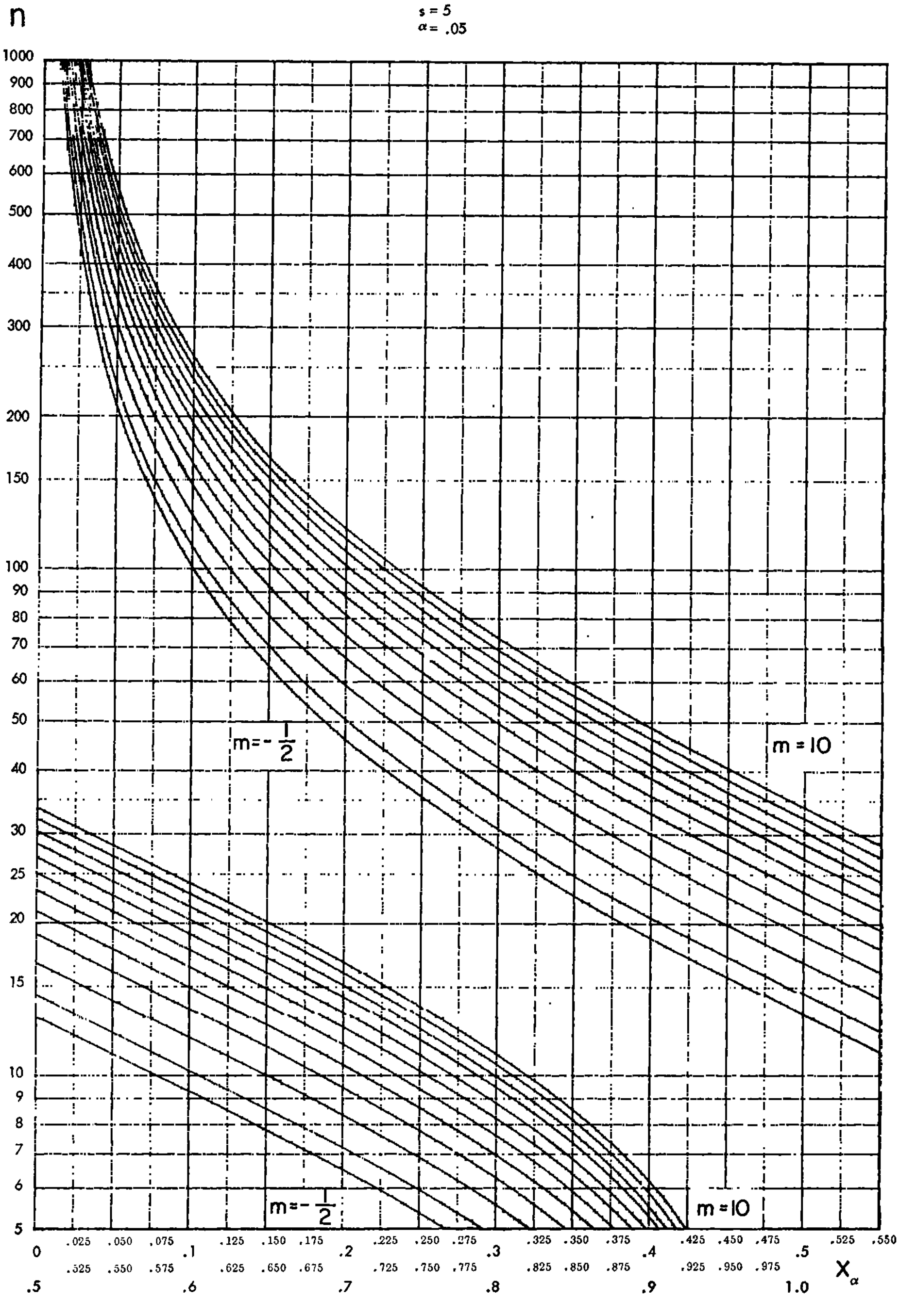

s = 5
α = .025
n
1000
900
800
700
600
500
400
300
200
150
100
90
80
70
60
50
40
30
25
20
15
10
9
8
7
6
5
m = -1/2
m = 10
m = -1/2
m = 10
0
.025
.050
.075
.1
.125
.150
.175
.2
.225
.250
.275
.3
.325
.350
.375
.4
.425
.450
.475
.5
.525
.550
.525
.550
.575
.625
.650
.675
.725
.750
.775
.825
.850
.875
.925
.950
.975
.5
.6
.7
.8
.9
1.0
Xα

CHART XII

s = 5
α = .05



3. Charts of the upper 1%, 2.5%, and 5% points of the distribution of the largest characteristic root.

3.1. *Description.* Charts I–XII enable finding $x_\alpha(s, m, n)$ such that

$$P[\theta_s \leqq x_\alpha(s, m, n)] = 1 - \alpha,$$

where θ_s is the largest non-zero root. On each page, the graphs appear for a particular s and α ($s = 2(1)5$, $\alpha = .01, .025, .05$) for $m = -\frac{1}{2}, 0(1)10$ and n from 5 to 1000. The curves corresponding to the twelve values of m on each page are in two sections, the lower section being the continuation of the upper section, with an overlap occurring from $x_\alpha = .50$ to $.55$. Of the two scales for x_α at the bottom of the page, the upper scale corresponds to the upper set of curves and the lower scale to the lower set. The lowest curve in each case (with the excep-

TABLE 4.1

Values of $z_\alpha(s, m)$

α / m	$s = 2$			$s = 3$		
	.01	.025	.05	.01	.025	.05
$-\frac{1}{2}$	12.1601	10.1465	8.5941	17.1762	14.9006	13.1141
0	14.5680	12.4157	10.7393	19.5012	17.1192	15.2389
1	18.7346	16.3599	14.4873	23.6906	21.1262	19.0866
2	22.4664	19.9086	17.8762	27.5181	24.7971	22.6216
3	25.9526	23.2352	21.0641	31.1203	28.2597	25.9635
4	29.2755	26.4145	24.1192	34.5647	31.5768	29.1708
5	32.4795	29.4870	27.0779	37.8905	34.7848	32.2774
6	35.5920	32.4773	29.9628	41.1230	37.9071	35.3050
7	38.6311	35.4018	32.7886	44.2795	40.9597	38.2685
8	41.6098	38.2722	35.5658	47.3726	43.9542	41.1785
9	44.5375	41.0970	38.3021	50.4118	46.8993	44.0430
10	47.4215	43.8827	41.0033	53.4042	49.8017	46.8684
α / m	$s = 4$			$s = 5$		
	.01	.025	.05	.01	.025	.05
$-\frac{1}{2}$	21.0646	19.4847	17.5183	26.6206	23.9697	21.8538
0	24.2395	21.6713	19.6277	28.8613	26.1339	23.9515
1	28.4328	25.7078	23.5278	33.0524	30.1861	27.8835
2	32.3175	29.4540	27.1543	36.9748	33.9834	31.5731
3	35.9964	33.0074	30.5996	40.7087	37.6027	35.0938
4	39.5253	36.4207	33.9135	44.3009	41.0883	38.4880
5	42.9387	39.7262	37.1265	47.7814	44.4688	41.7829
6	46.2593	42.9454	40.2588	51.1710	47.7639	44.9971
7	49.5034	46.0934	43.3246	54.4847	50.9876	48.1441
8	52.6831	49.1815	46.3345	57.7338	54.1508	51.2340
9	55.8073	52.2182	49.2964	60.9269	57.2615	54.2745
10	58.8833	55.2102	52.2166	64.0709	60.3264	57.2717

tion of Chart III) corresponds to $m = -\frac{1}{2}$, the next lowest to $m = 0$, the next to $m = 1$, etc., to the uppermost curve, which corresponds to $m = 10$. The scale for n is on the left margin of the page and is logarithmic.

3.2. *Note.* The values of $x_\alpha(s, m, n)$ may be read from the charts correct to two decimals. For a more precise value, when $n > 100$, the method described in Section 4 is suggested.

4. Asymptotic $z_\alpha(s, m)$ values.

4.1. *Description.* In Table 4.1 the values of $z_\alpha(s, m)$ are listed for $s = 2(1)5$, $m = -\frac{1}{2}, 0(1)10$, and $\alpha = .01, .025, .05$. For $n > 100$, these may be used to obtain $x_\alpha(s, m, n)$, with an error of at most five units in the fourth decimal. For a given combination (s, m, n) and a desired significance level α, determine $x = x_\alpha(s, m, n)$ from (2.1) with $z = z_\alpha(s, m)$ obtained from Table 4.1.

5. Acknowledgments. I should like to express my sincere thanks to S. N. Roy and R. E. Bargmann for their helpful advice and assistance in the preparation of this paper.

REFERENCES

[1] Hebron E. Adams, "DOPSIR: Double precision floating point SOAP interpretive routine," 650 Program Library, File No. 2.0.010, IBM, Washington, D.C. (1956).

[2] T. W. Anderson, *Introduction to Multivariate Statistical Analysis*, John Wiley and Sons, New York, 1958.

[3] Rolf Bargmann, "A study of independence and dependence in multivariate normal analysis," North Carolina Institute of Statistics Mimeograph Series No. 186 (1957).

[4] M. S. Bartlett, "Further aspects of the theory of multiple regression," *Proc. Cambridge Phil. Soc.*, Vol. 34 (1938), pp. 33–40.

[5] S. B. Chaudhuri, "Statistical tables and certain recurrence relations connected with p-statistics," *Calcutta Stat. Assoc. Bull.*, Vol. 6 (1956), pp. 181–188.

[6] R. A. Fisher, "The sampling distribution of some statistics obtained from non-linear equations," *Ann. Eugenics*, Vol. 9 (1939), pp. 238–249.

[7] F. G. Foster and D. H. Rees, "Upper percentage points of the generalized beta distribution. I," *Biometrika*, Vol. 44 (1957), pp. 237–247.

[8] F. G. Foster, "Upper percentage points of the generalized beta distribution. II," *Biometrika*, Vol. 44 (1957), pp. 441–453.

[9] F. G. Foster, "Upper percentage points of the generalized beta distribution. III," *Biometrika*, Vol. 45 (1958), pp. 492–503.

[10] M. A. Girshick, "On the sampling theory of the roots of determinantal equations," *Ann. Math. Stat.*, Vol. 10 (1939), pp. 203–224.

[11] P. L. Hsu, "On the distribution of roots of certain determinantal equations," *Ann. Eugenics*, Vol. 9 (1939), pp. 250–258.

[12] D. N. Nanda, "Distribution of a root of a determinantal equation," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 47–57.

[13] D. N. Nanda, "Probability distribution tables of the larger root of a determinantal equation with two roots," *J. Indian Soc. of Agricultural Stat.*, Vol. 3 (1951), pp. 175–177.

[14] K. C. Sreedharan Pillai, "On the distribution of the largest or the smallest root of a matrix in multivariate analysis," *Biometrika*, Vol. 43 (1956), pp. 122–127.

[15] K. C. Sreedharan Pillai, *Concise Tables for Statisticians*, The Statistical Center University of the Philippines, Manila, 1957.

[16] K. C. Sreedharan Pillai and Celia G. Bantegui, "On the distribution of the largest of six roots of a matrix in multivariate analysis," *Biometrika*, Vol. 46 (1959), pp. 237–240.

[17] C. Radhakrishna Rao, *Advanced Statistical Methods in Biometric Research*, John Wiley and Sons, New York, 1952.

[18] S. N. Roy, "p-statistics or some generalizations in analysis of variance appropriate to multivariate problems," *Sankhyā*, Vol. 4 (1939), pp. 381–396.

[19] Samarendra Nath Roy, "The individual sampling distribution of the maximum, the minimum and any intermediate of the p-statistics on the null hypothesis," *Sankhyā*, Vol. 7 (1945), pp. 133–158.

[20] S. N. Roy, "On a heuristic method of test construction and its use in multivariate analysis," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 220–238.

[21] S. N. Roy and R. C. Bose, "Simultaneous confidence interval estimation," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 513–536.

[22] S. N. Roy, "Some further results in simultaneous confidence interval estimation," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 752–761.

[23] S. N. Roy, *Some Aspects of Multivariate Analysis*, John Wiley and Sons, New York, 1957.

[24] S. N. Roy and R. Gnanadesikan, "Further contributions to multivariate confidence bounds," *Biometrika*, Vol. 44 (1957), pp. 399–410.

[25] S. N. Roy and R. E. Bargmann, "Tests of multiple independence and the associated confidence bounds," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 491–503.

[26] S. N. Roy and R. F. Potthoff, "Confidence bounds on vector analogues of the 'ratio of means' and the 'ratio of variances' for two correlated normal variates and some associated tests," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 829–841.

[27] S. N. Roy and J. Roy, "Analysis of variance with univariate or multivariate, fixed or mixed classical models," North Carolina Institute of Statistics Mimeograph Series No. 208 (1958).

[28] John R. B. Whittlesey, "Summary of results obtained under the Ford Foundation contract," unpublished research report, Dept. of Statistics, University of North Carolina (1954).

[29] S. S. Wilks, "Certain generalizations in the analysis of variance," *Biometrika*, Vol. 24 (1932), pp. 471–494.

[30] S. S. Wilks, *Mathematical Statistics*, Princeton, N. J., Princeton Univ. Press, 1943.

[31] V. M. Wolontis, "A complete floating-decimal interpretive system for the IBM 650 magnetic drum calculator," IBM Applied Science Div., *Tech. Newsletter No. 11* (1956).

CHART I

$s = 2$
$\alpha = .01$

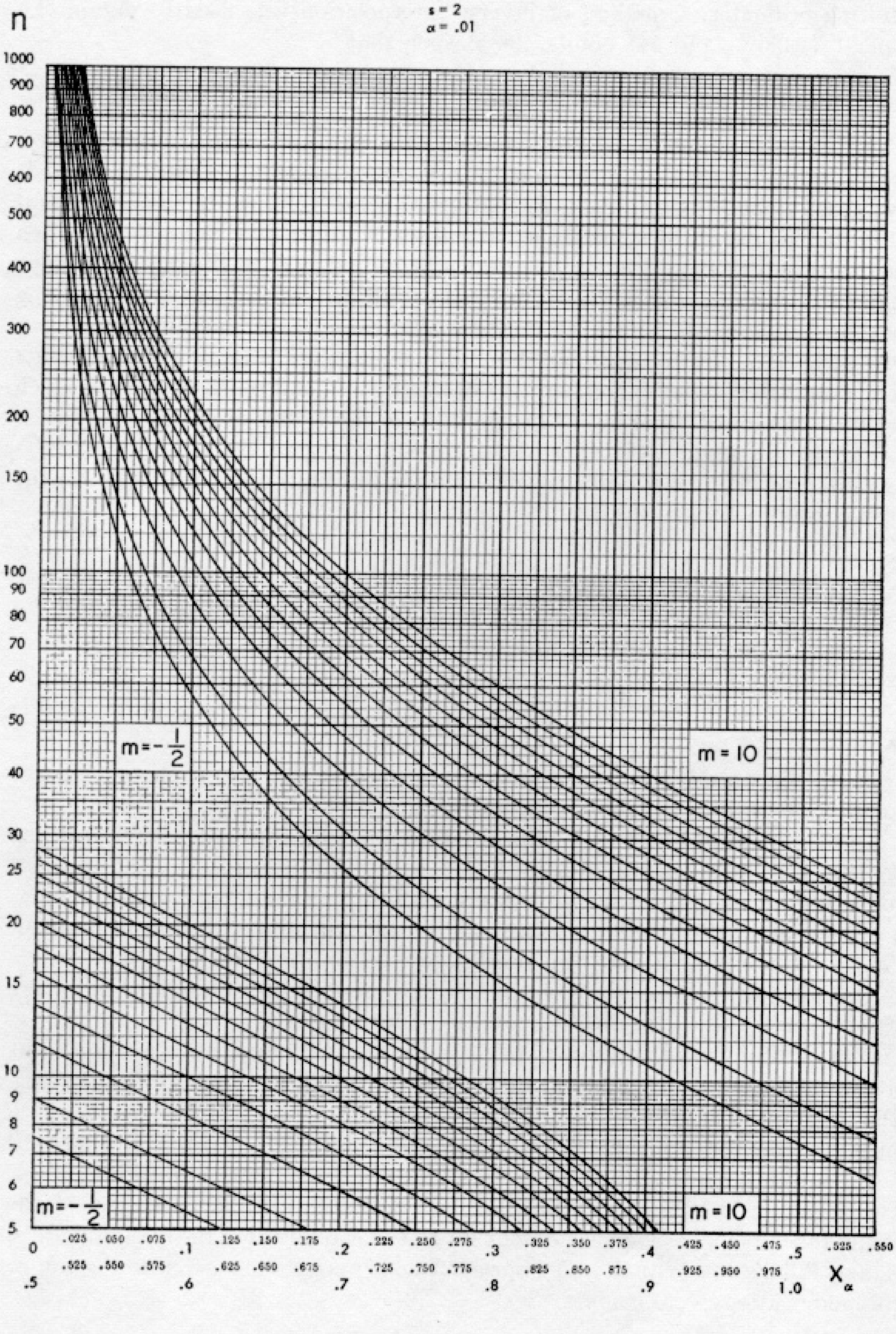


CHART II

$s = 2$

$\alpha = .025$

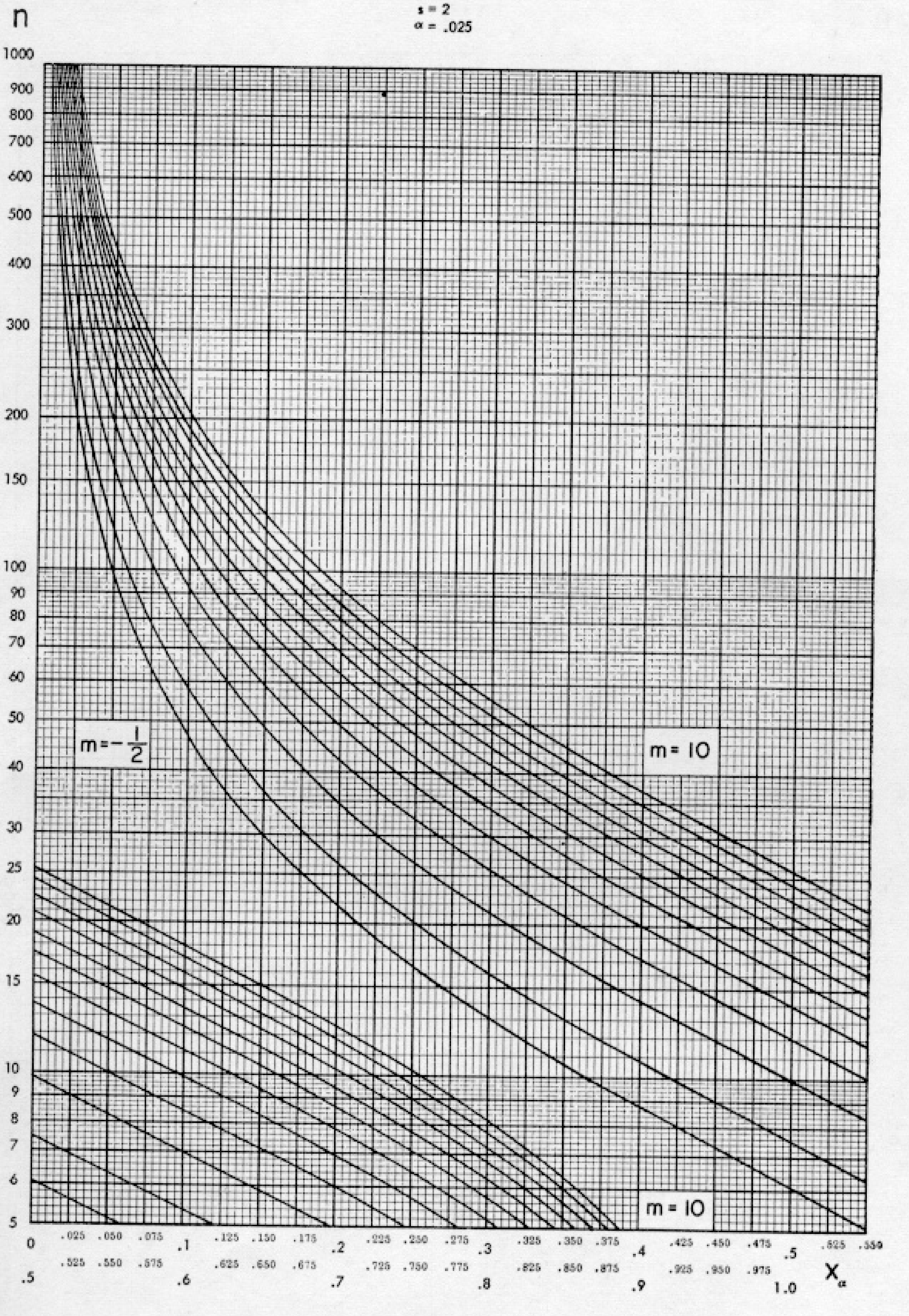


CHART III

s = 2
α = .05

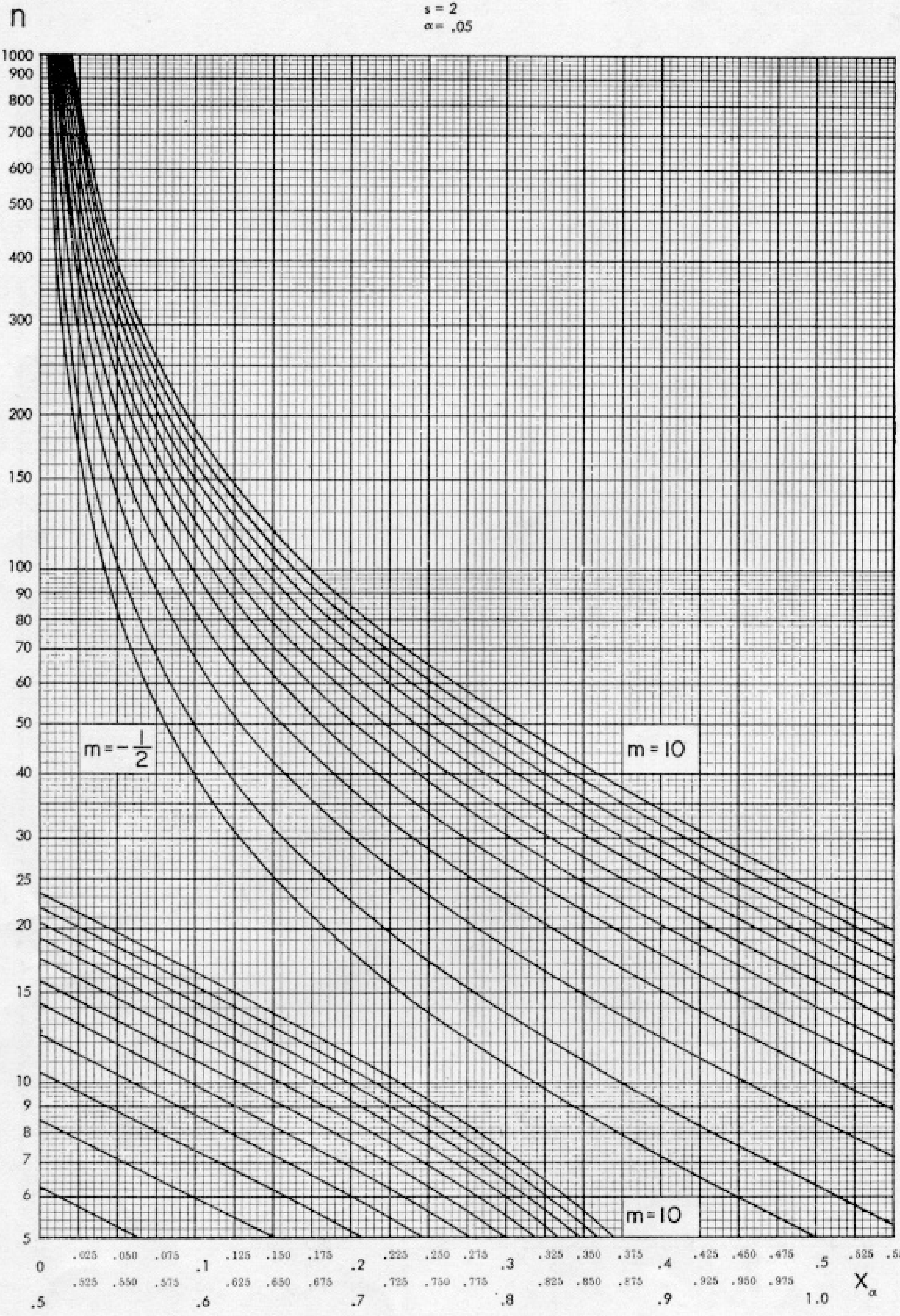


CHART VII

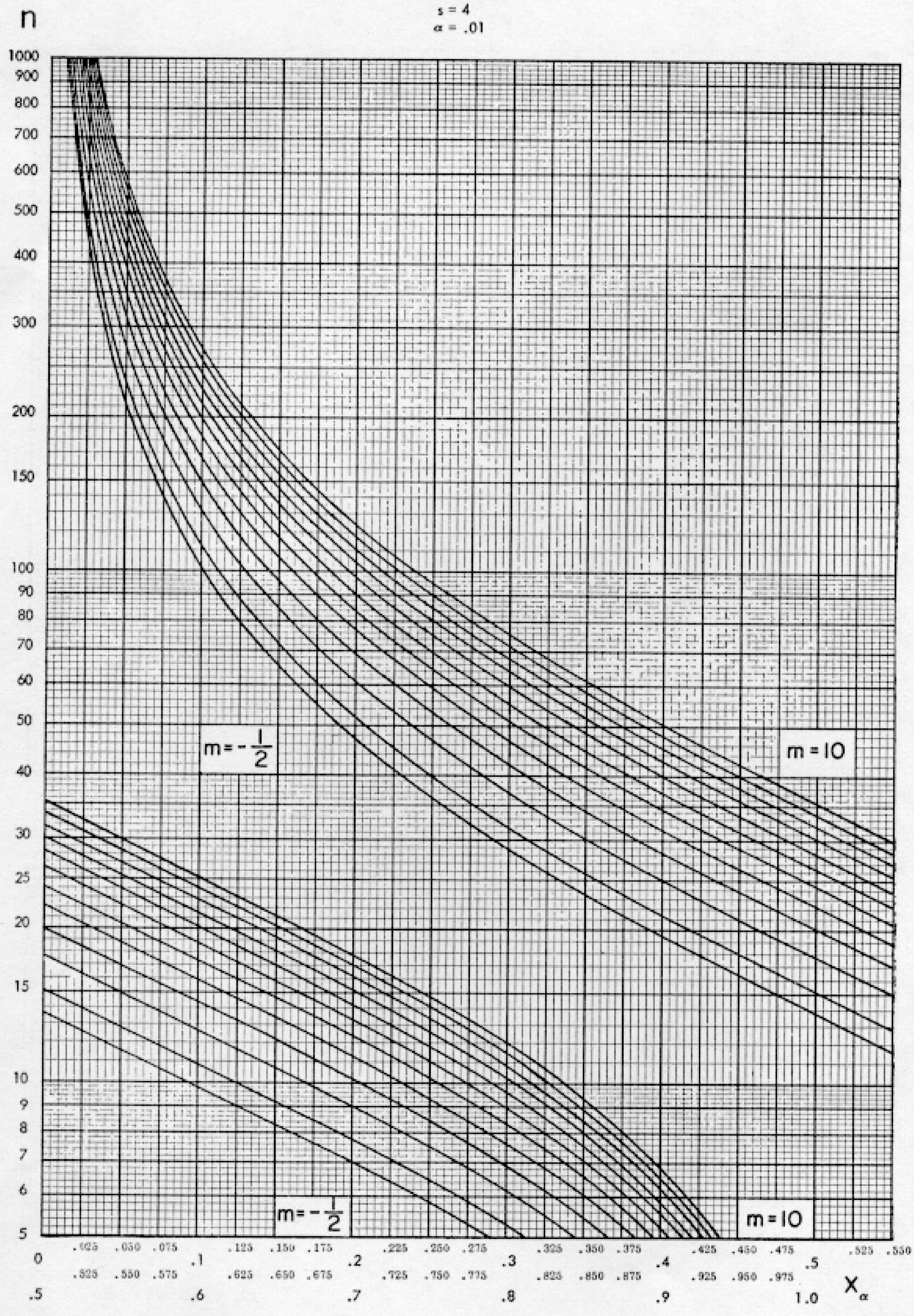


CHART VIII

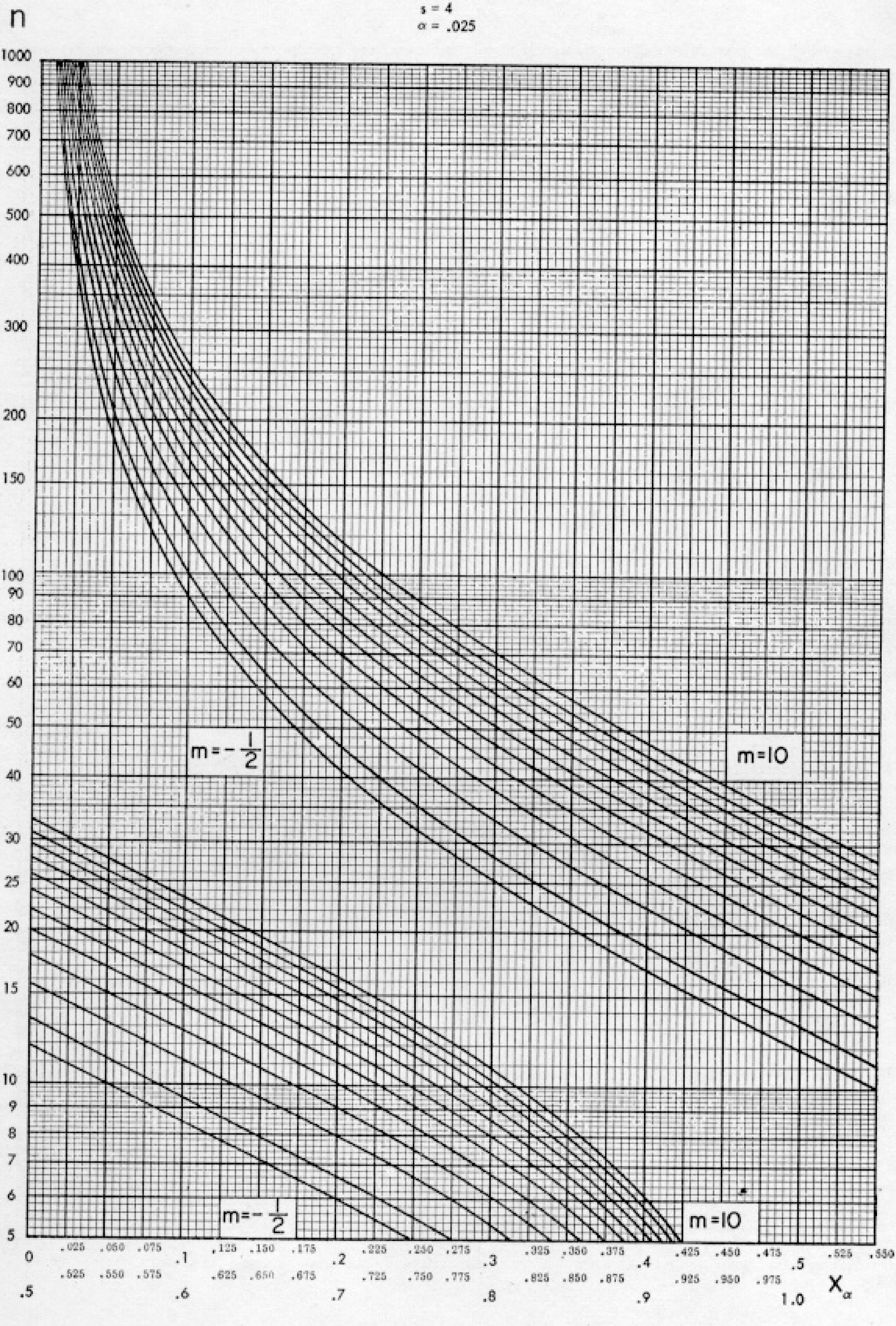

s = 4
α = .025
n
1000
900
800
700
600
500
400
300
200
150
100
90
80
70
60
50
40
30
25
20
15
10
9
8
7
6
5
m = − 1/2
m = 10
m = − 1/2
m = 10
0
.025
.050
.075
.1
.125
.150
.175
.2
.225
.250
.275
.3
.325
.350
.375
.4
.425
.450
.475
.5
.525
.550
.5
.525
.550
.575
.6
.625
.650
.675
.7
.725
.750
.775
.8
.825
.850
.875
.9
.925
.950
.975
1.0
X_α

CHART IX

$s = 4$
$\alpha = .05$

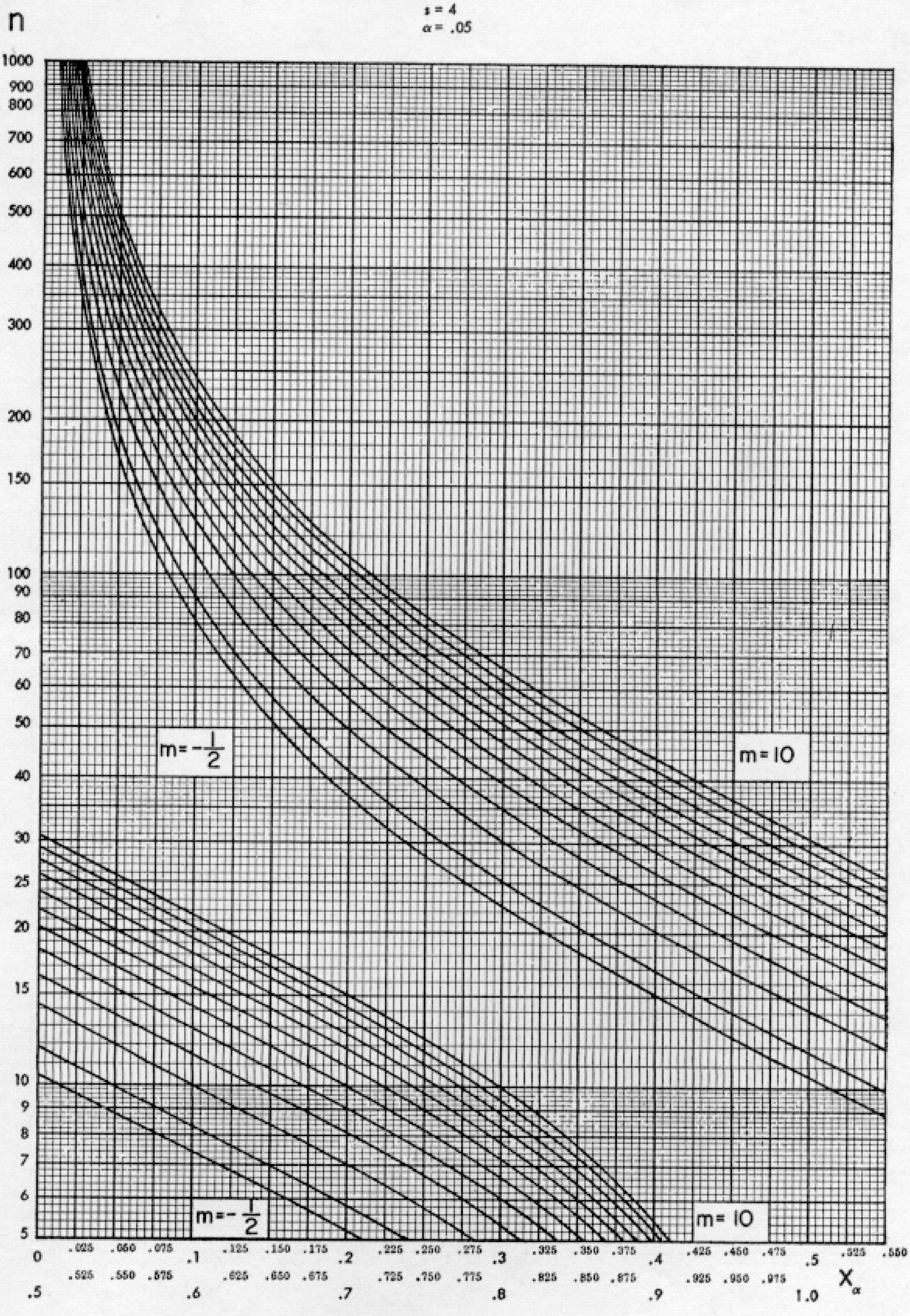


CHART X

$s = 5$
$\alpha = .01$

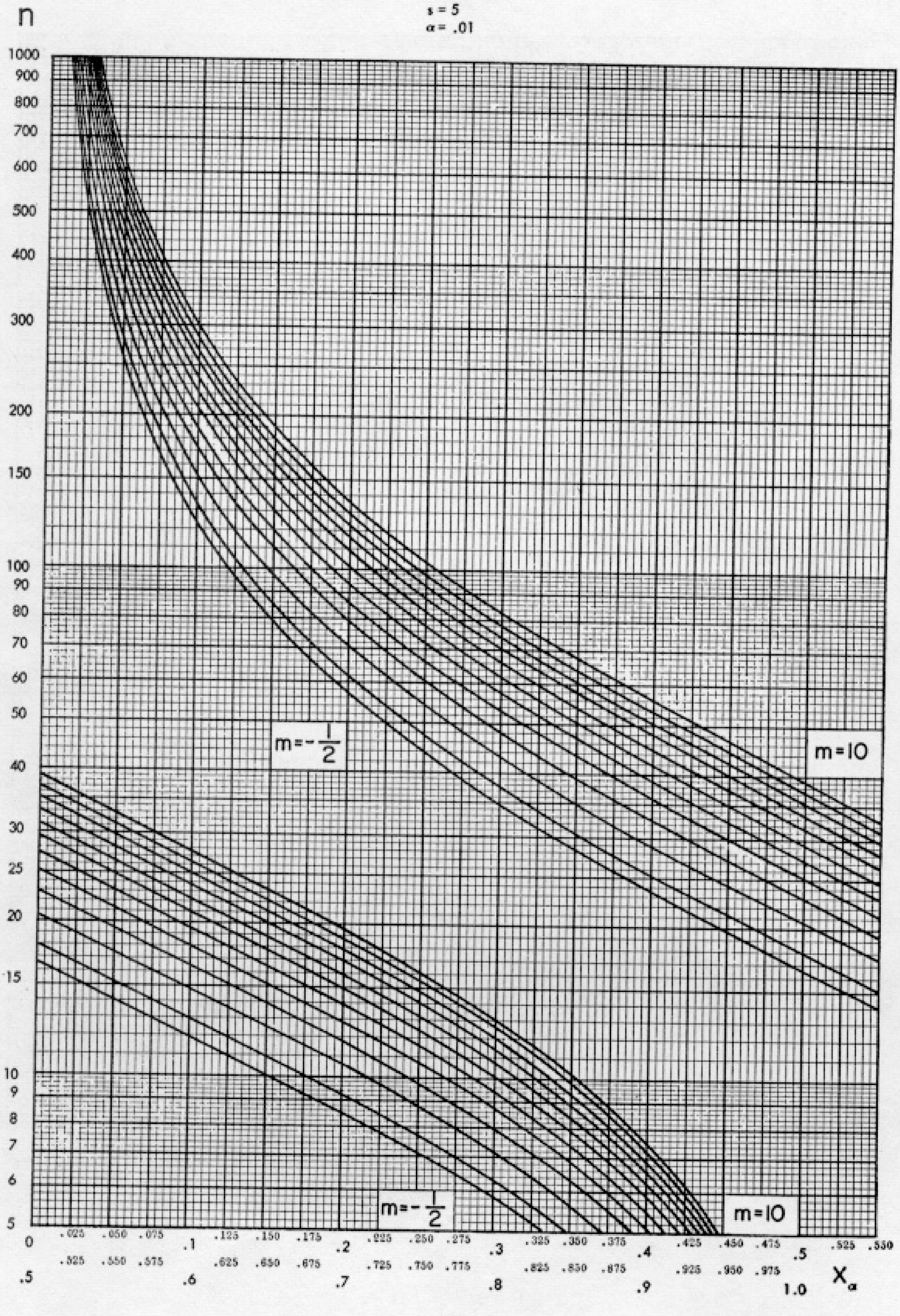


CHART XI

s = 5
α = .025

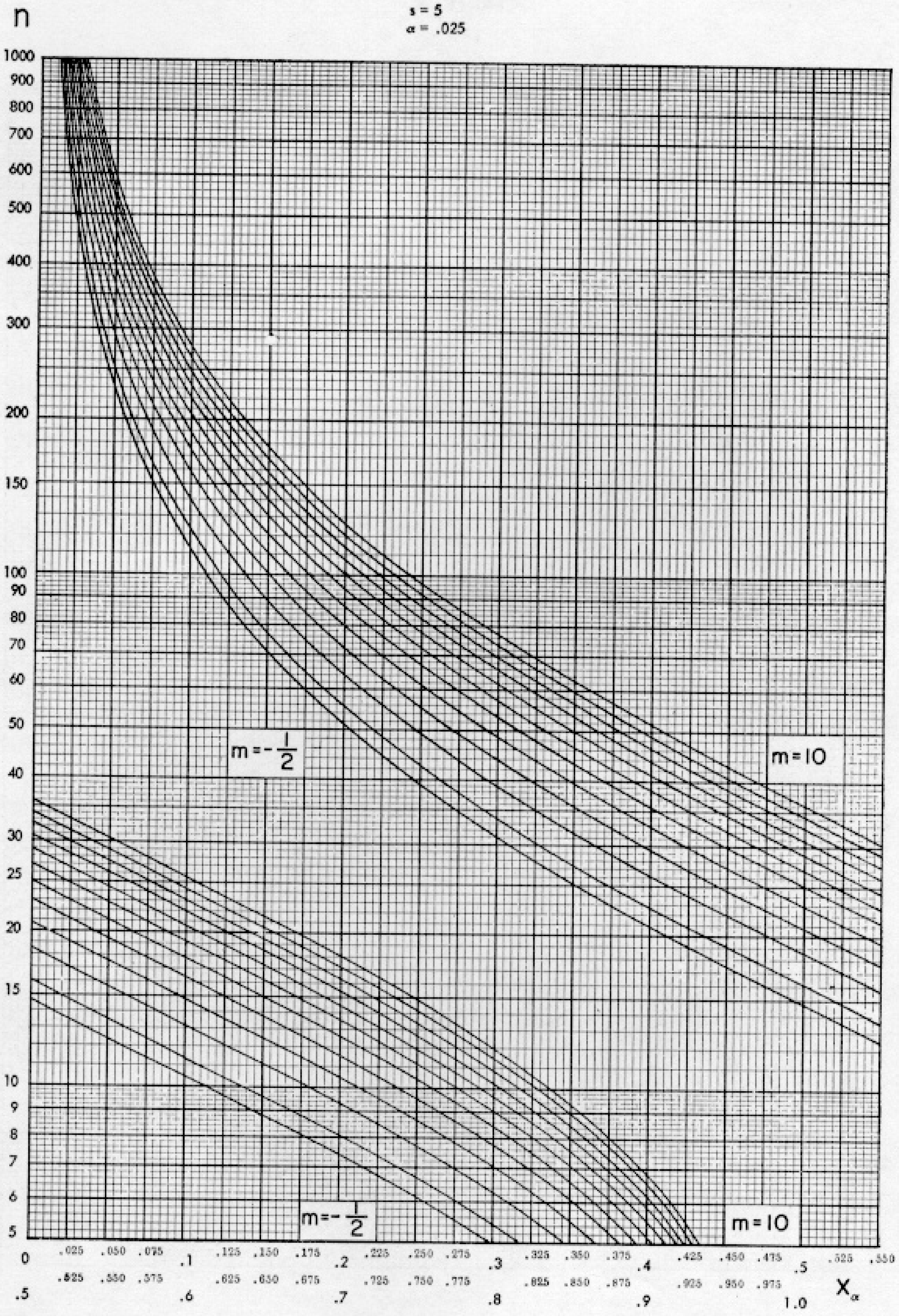
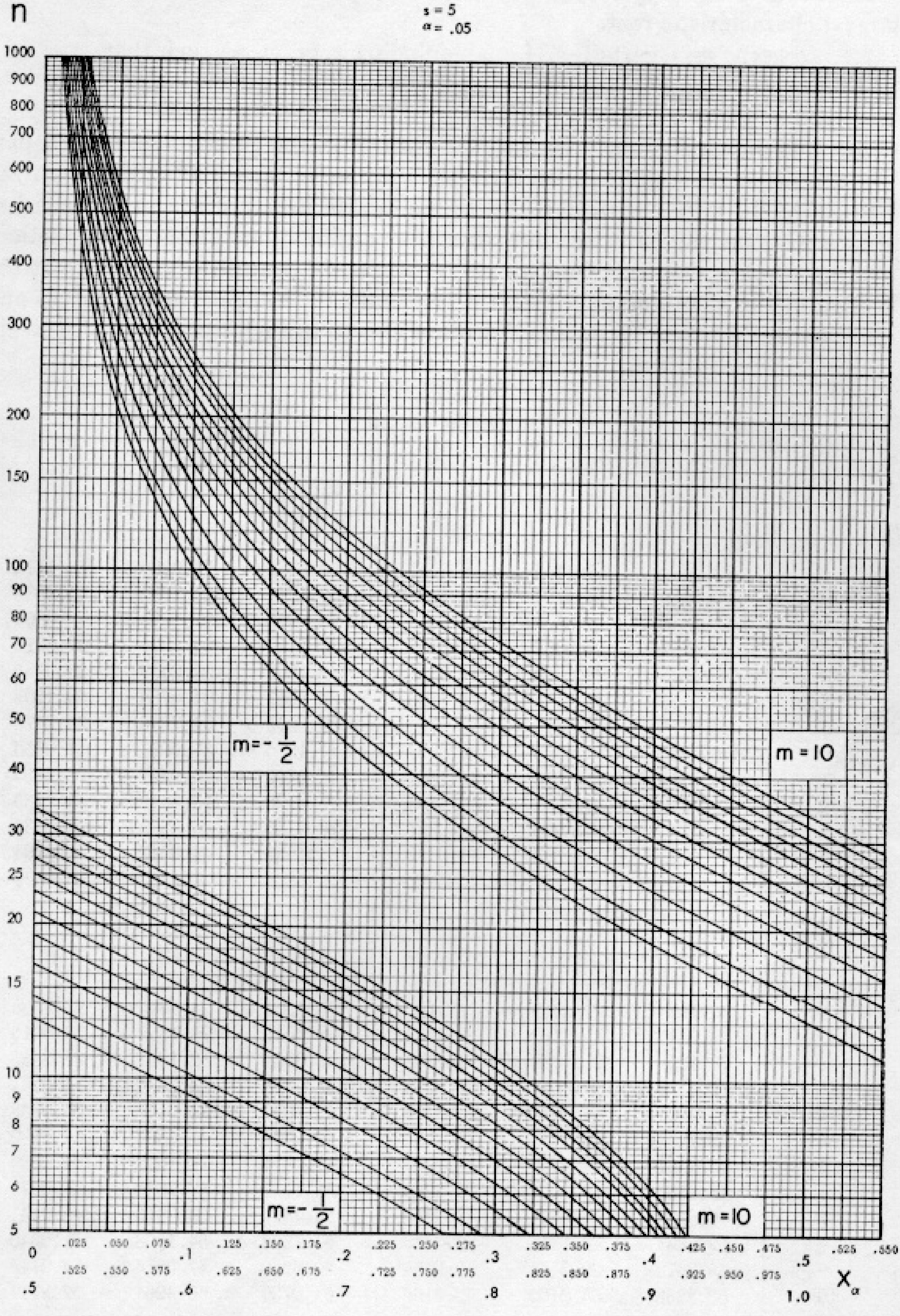


CHART XII


s = 5
α = .05
n
1000
900
800
700
600
500
400
300
200
150
100
90
80
70
60
50
40
30
25
20
15
10
9
8
7
6
5
m = - 1/2
m = 10
m = - 1/2
m = 10
0
.025
.050
.075
.1
.125
.150
.175
.2
.225
.250
.275
.3
.325
.350
.375
.4
.425
.450
.475
.5
.525
.550
.5
.525
.550
.575
.6
.625
.650
.675
.7
.725
.750
.775
.8
.825
.850
.875
.9
.925
.950
.975
1.0
X_α