TABLES OF PERCENTILE POINTS OF A VECTOR IN PRINCIPAL COMPONENT ANALYSIS.

by

T. Sugiyama

University of North Carolina

Institute of Statistics Mimeo Series No. 582.

May 1968

This research was supported by the National Science Foundation Grant No. GU-2059 and the Air Force Office of Scientific Research Grant No. AF-AFOSR-68-1415 and the National Institute of Health Grant No. 5401-GM12868-04.

DEPARTMENT OF STATISTICS

University of North Carolina

Chapel Hill, N. C.

INTRODUCTION

Let $X_1, \ldots, X_N$ be a random sample from bivariate normal distribution $N(\mu,\Sigma)$. Then the 2×2 positive definite symmetric matrix $U = \Sigma_{\alpha=1}^{N} (X_\alpha-\bar{X})(X_\alpha-\bar{X})'$ can be diagonalized as follows

$$\begin{pmatrix} U_{11} & U_{12} \\ U_{12} & U_{22} \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

where $\infty > \lambda_1 > \lambda_2 > 0$, and $-\frac{\pi}{2} < \theta < +\frac{\pi}{2}$. The vectors $(\cos\theta, \sin\theta)'$ and $(-\sin\theta, \cos\theta)'$ may be called the latent vectors corresponding to the first principal component, $(\cos\theta)\, x_1 + (\sin\theta)x_2$, and the second principal component, $-(\sin\theta)x_1 + (\cos\theta)x_2$, respectively. Then we know the probability density function of the variate θ is

$$f(\theta;\Sigma) = \frac{1}{\pi(n+1)} \left(\frac{\sigma_1\sigma_2}{\bar{\sigma}^2}\right)^{n/2}$$

$$\left\{\frac{n+1}{2}\, {}_2F_1\left(1,n; \frac{n+1}{2}; x\right) - \frac{n-1}{2}\, {}_2F_1\left(1,n; \frac{n+3}{2}; x\right)\right\}, \quad \left(-\frac{\pi}{2} < \theta < \frac{\pi}{2}\right)$$

where σ_1 and σ_2 $(\sigma_1 \geq \sigma_2 > 0)$ are the latent roots of the population covariance matrix Σ, $n = N - 1$,

$$x = (\sigma_{11} \cos^2\theta + 2\sigma_{12} \sin\theta \cos\theta + \sigma_{22} \sin^2\theta)/(\sigma_{11} + \sigma_{22}),$$

and $\bar{\sigma} = (\sigma_1 + \sigma_2)/2$. If the matrix $\Sigma = I$, the distribution of the variate θ is the uniform distribution, namely $f(\theta) = 1/\pi$. For the population covariance matrix Σ, let us consider the diagonal matrix $\Sigma = \mathrm{diag}\,(\sigma_1,\sigma_2)$ and mean vector $\mu = 0$ without the loss of generality, where $\sigma_1 \geq \sigma_2 > 0$. Then we notice that the population latent vectors are the vector $(1\ 0)'$ corresponding to the population largest root σ_1 and the vector $(0\ 1)'$ the root σ_2 respectively, the above $x = \sigma_1 \cos^2\theta + \sigma_2 \sin^2\theta$, and the symmetricity of the probability density function $f(\theta) = f(-\theta)$. The distribution gives us information on the stability of the principal axis corresponding to the largest latent root λ_1, and also the latent root λ_2 in the case $p = 2$.

Figures 1 and 2 show for different values of $\sigma_1(\sigma_1 + \sigma_2 = 2)$ the curves of the points x satisfying $P(x < \theta < \pi/2) = 0.005$ and $P(x < \theta < \pi/2) = 0.025$ respectively, corresponding to different values of n.

$$\int_{x}^{\pi/2} f(\theta;\Sigma)\, d\theta = 0.005$$

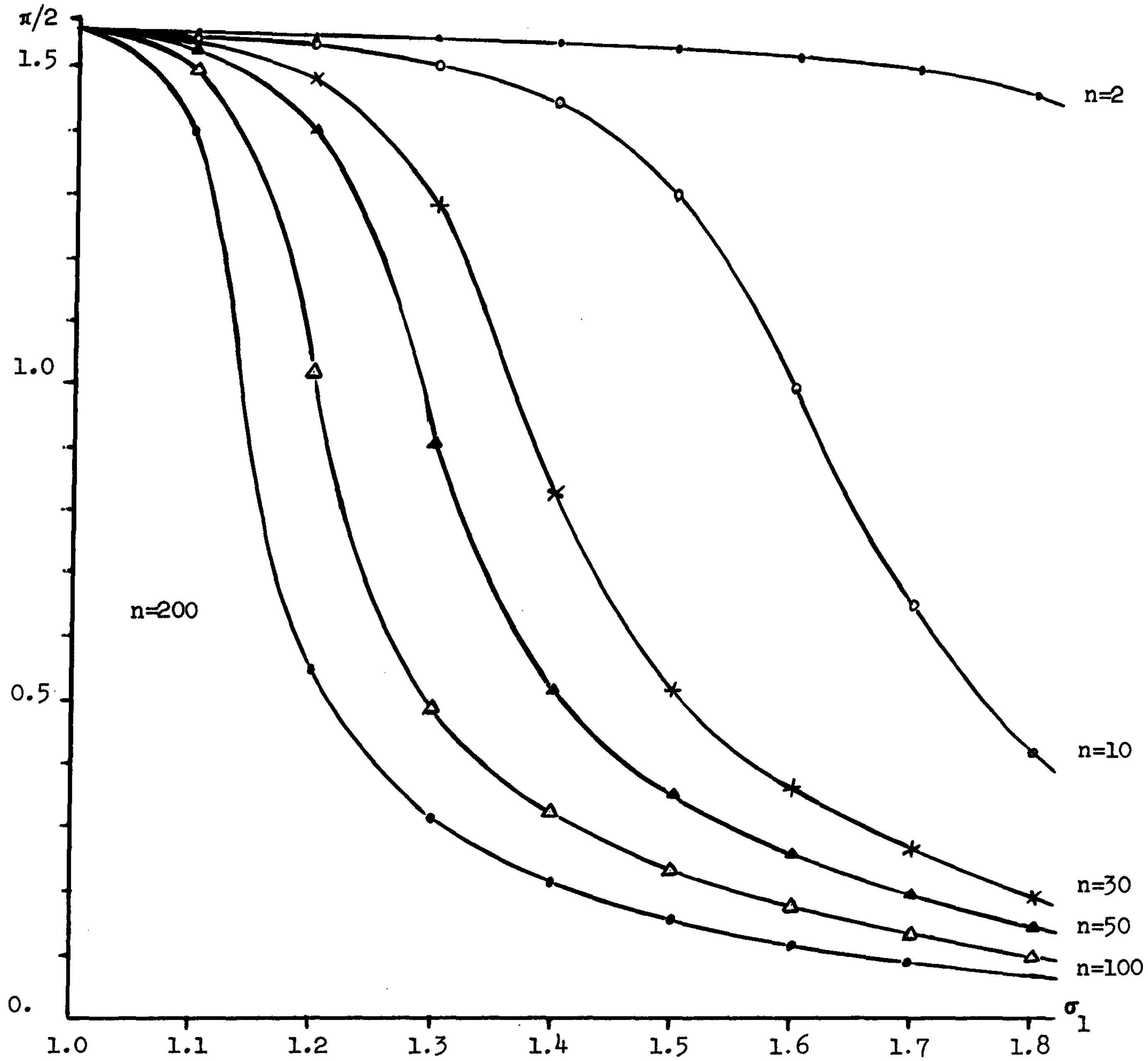

Figure 1

$$\int_x^{\pi/2} f(\theta;\Sigma)\, d\theta = 0.025$$

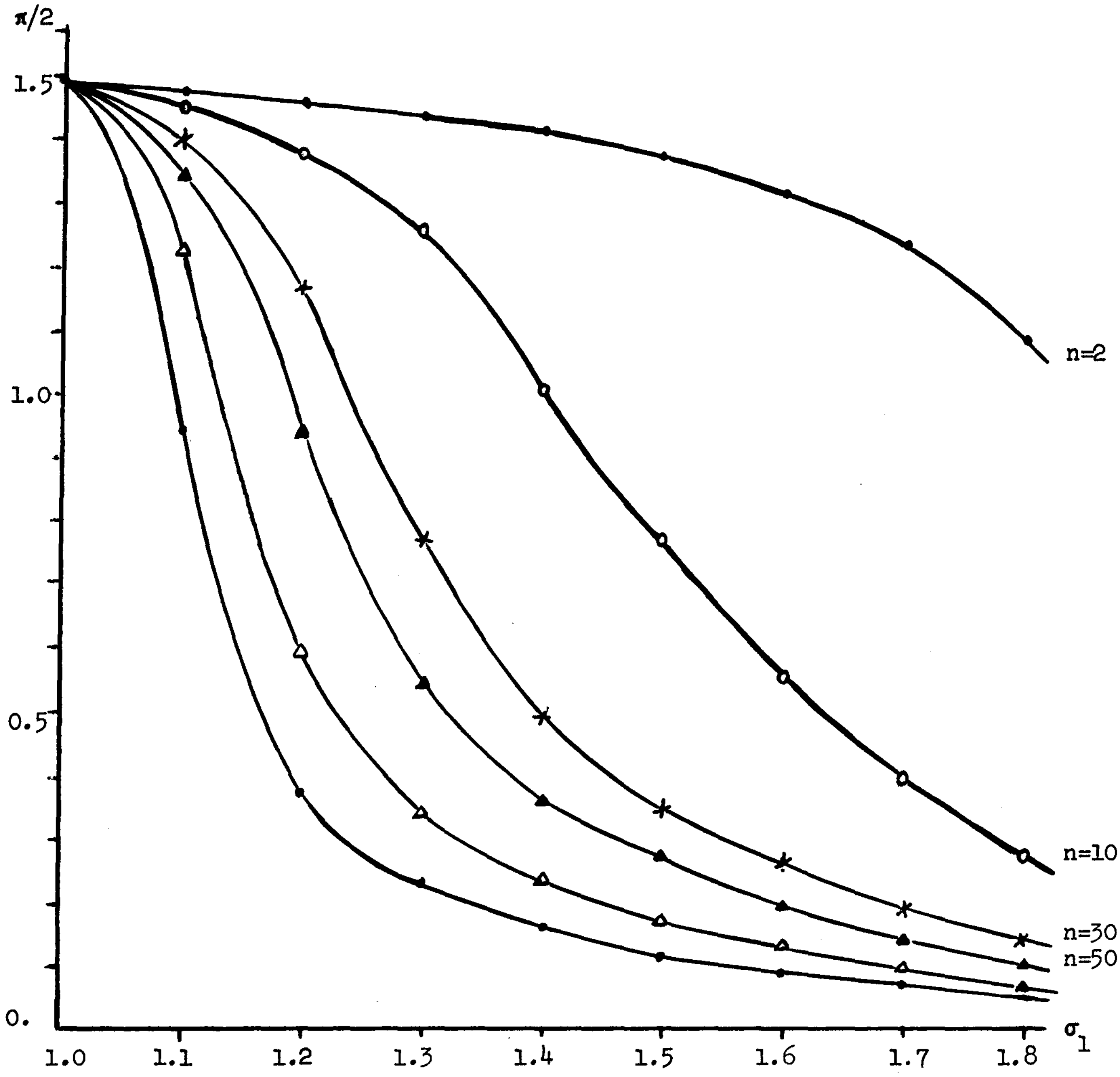

Figure 2

Tables 1 and 2 give for different values of σ_1 $(\sigma_1 + \sigma_2 = 2)$ the points x such that $P(x < \theta < \pi/2) = 0.005$ and $P(x < \theta < \frac{\pi}{2}) = 0.025$, corresponding to 40 different values of n.

Calculations were carried out on IBM 360 model 75 machine using double precision arithmetic by trapezoidal rule for each interval the ranges from 0 to $\pi/2$ dividing 1500 intervals. The convergence of hypergeometric functions was checked evaluating the kth term

$$a_k = ((n)_k/((n+1)/2)_k)((\sigma_1 \cos^2\theta + \sigma_2 \sin^2\theta)/2)^k,$$

namely, $a_k \leq 10^{-30}$ for the values n from 2 to 50, $a_k \leq 10^{-20}$ from 55 to 140, and so on. The values shown to 5 decimal places were based on original calculations which were believed to be correct to about 7 decimal points for $P(x < \theta < \pi/2) = 0.025$ and 8 decimal points for $P(x < \theta < \pi/2) = 0.005$.

ACKNOWLEDGEMENT; The author wishes to express his sincere thanks to Mr. James O. Kitchen for his valuable suggestions in the computation, and Professor Norman L. Johnson for his kind encouragement.

REFERENCE

[1] Sugiyama, T. (1965). On the distribution of the latent vectors for principal component analysis. Ann. Math. Statist. 36, 1875-1876.

Table 1.

$$\int_x^{\pi/2} f(\theta,\Sigma)\, d\theta = 0.005$$

$(\sigma_1 + \sigma_2 = 2)$

n \ σ_1	1.1	1.2	1.3	1.4
2	1.55236	1.54900	1.54470	1.53901
4	1.55074	1.54465	1.53573	1.52192
6	1.54940	1.54062	1.52629	1.50115
8	1.54819	1.53665	1.51592	1.47534
10	1.54705	1.53261	1.50435	1.44329
12	1.54597	1.52845	1.49137	1.40384
14	1.54491	1.52413	1.47676	1.35607
16	1.54388	1.51964	1.46034	1.29960
18	1.54287	1.51493	1.44192	1.23501
20	1.54187	1.50999	1.42132	1.16418
22	1.54087	1.50481	1.39837	1.09016
24	1.53988	1.49937	1.37296	1.01654
26	1.53889	1.49364	1.34501	0.94654
28	1.53790	1.48762	1.31451	0.88229
30	1.53691	1.48128	1.28157	0.82477
32	1.53592	1.47462	1.24638	0.77402
34	1.53492	1.46761	1.20926	0.72955
36	1.53391	1.46024	1.17063	0.69063
38	1.53291	1.45249	1.13103	0.65649
40	1.53189	1.44435	1.09101	0.62642
42	1.53087	1.43581	1.05116	0.59980
44	1.52984	1.42684	1.01201	0.57609
46	1.52880	1.41745	0.97403	0.55486
48	1.52775	1.40761	0.93757	0.53573
50	1.52669	1.39732	0.90290	0.51841
55	1.52399	1.36955	0.82477	0.48145
60	1.52123	1.33880	0.75883	0.45142
65	1.51839	1.30511	0.70369	0.42644
70	1.51246	1.26863	0.65752	0.40525
75	1.51247	1.22967	0.61856	0.38698

Table 1 continued

n \ σ_1	1.1	1.2	1.3	1.4
80	1.50938	1.18869	0.58536	0.37103
85	1.50619	1.14631	0.55677	0.35693
90	1.50290	1.10323	0.53189	0.34435
95	1.49951	1.06018	0.51002	0.33303
100	1.49602	1.01786	0.49064	0.32278
110	1.48869	0.93763	0.45774	0.30487
120	1.48090	0.86563	0.43075	0.28967
140	1.46378	0.74890	0.38877	0.26509
160	1.44442	0.66305	0.35727	0.24590
200	1.39814	0.54987	0.31228	0.21747

n \ σ_1	1.5	1.6	1.7	1.8
2	1.53109	1.51927	1.49967	1.46081
4	1.49897	1.45683	1.36822	1.15168
6	1.45282	1.34951	1.11826	0.74185
8	1.38747	1.18726	0.84081	0.52815
10	1.29906	0.99608	0.65367	0.42270
12	1.18876	0.82730	0.54113	0.36081
14	1.06638	0.70218	0.46894	0.31958
16	0.94723	0.61319	0.41882	0.28975
18	0.84294	0.54855	0.38176	0.26691
20	0.75717	0.49985	0.35302	0.24872
22	0.68823	0.46183	0.32992	0.23379
24	0.63278	0.43125	0.31085	0.22125
26	0.58769	0.40601	0.29475	0.21052
28	0.55045	0.38476	0.28092	0.20122
30	0.51922	0.36655	0.26888	0.19304
32	0.49263	0.35072	0.25827	0.18578
34	0.46970	0.33680	0.24883	0.17929
36	0.44969	0.32443	0.24035	0.17342
38	0.43204	0.31335	0.23269	0.16810
40	0.41634	0.30334	0.22572	0.16324

Table 1 continued

n \ σ_1	1.5	1.6	1.7	1.8
42	0.40225	0.29423	0.21933	0.15878
44	0.38952	0.28591	0.21346	0.15466
46	0.37794	0.27826	0.20804	0.15085
48	0.36735	0.27120	0.20302	0.14730
50	0.35762	0.26465	0.19834	0.14399
55	0.33636	0.25016	0.18792	0.13661
60	0.31853	0.23783	0.17899	0.13026
65	0.30330	0.22716	0.17122	0.12472
70	0.29008	0.21781	0.16439	0.11983
75	0.27847	0.20954	0.15831	0.11547
80	0.26816	0.20214	0.15286	0.11155
85	0.25893	0.19547	0.14794	0.10801
90	0.25059	0.18943	0.14346	0.10479
95	0.24301	0.18391	0.13937	0.10184
100	0.23608	0.17885	0.13560	0.09912
110	0.22385	0.16987	0.12891	0.09428
120	0.21334	0.16212	0.12312	0.09008
140	0.19613	0.14935	0.11356	0.08315
160	0.18251	0.13919	0.10592	0.07760
200	0.16208	0.12386	0.09436	

Table 2

$$\int_x^{\pi/2} f(\theta,\Sigma)\ d\theta = 0.025$$

$$(\sigma_1 + \sigma_2 = 2)$$

n \ σ_1	1.1	1.2	1.3	1.4
2	1.47871	1.46201	1.44083	1.41302
4	1.47067	1.44062	1.39727	1.33231
6	1.46402	1.42097	1.35270	1.24175
8	1.45804	1.40175	1.30558	1.14333
10	1.45244	1.38247	1.25571	1.04285
12	1.44710	1.36289	1.20347	0.94697
14	1.44193	1.34290	1.14963	0.86052
16	1.43689	1.32243	1.09521	0.78578
18	1.43194	1.30147	1.04130	0.72166
20	1.42705	1.28000	0.98892	0.66778
22	1.42221	1.25806	0.93891	0.62223
24	1.41740	1.23569	0.89184	0.58351
26	1.41262	1.21294	0.84805	0.55034
28	1.40784	1.18987	0.80765	0.52166
30	1.40307	1.16657	0.77058	0.49666
32	1.39831	1.14312	0.73669	0.47468
34	1.39354	1.11959	0.70575	0.45520
36	1.38875	1.09609	0.67751	0.43782
38	1.38396	1.07270	0.65172	0.42221
40	1.37915	1.04950	0.62813	0.40810
42	1.37432	1.02659	0.60651	0.39527
44	1.36947	1.00402	0.58664	0.38355
46	1.36460	0.98187	0.56836	0.37280
48	1.35971	0.96019	0.55147	0.36289
50	1.35479	0.93903	0.53585	0.35373
55	1.34237	0.88866	0.50147	0.33352
60	1.32978	0.84214	0.47254	0.31641
65	1.31701	0.79956	0.44784	0.30170
70	1.30404	0.76082	0.42650	0.28886
75	1.29090	0.72568	0.40785	0.27754

Table 2 continued

n \ σ_1	1.1	1.2	1.3	1.4
80	1.27757	0.69383	0.39140	0.26746
85	1.26407	0.66496	0.37675	0.25841
90	1.25040	0.63875	0.36362	0.25022
95	1.23658	0.61491	0.35176	0.24277
100	1.22262	0.59315	0.34098	0.23595
110	1.19433	0.55500	0.32210	0.22388
120	1.16568	0.52271	0.30604	0.21349
140	1.10790	0.47109	0.28005	0.19644
160	1.05057	0.43166	0.25975	0.18292
200	0.94197	0.37507	0.22968	0.16259

n \ σ_1	1.5	1.6	1.7	1.8
2	1.37485	1.31936	1.23222	1.08059
4	1.23186	1.07681	0.85813	0.60627
6	1.06774	0.83723	0.60609	0.41455
8	0.90951	0.66648	0.47361	0.32771
10	0.77839	0.55568	0.39692	0.27815
12	0.67770	0.48148	0.34711	0.24557
14	0.60158	0.42893	0.31190	0.22219
16	0.54329	0.38978	0.28548	0.20438
18	0.49763	0.35939	0.26475	0.19025
20	0.46101	0.33502	0.24795	0.17869
22	0.43100	0.31495	0.23398	0.16901
24	0.40594	0.29808	0.22213	0.16074
26	0.38466	0.28365	0.21192	0.15357
28	0.36633	0.27113	0.20300	0.14728
30	0.35036	0.26014	0.19511	0.14172

Table 2 continued

n \ σ_1	1.5	1.6	1.7	1.8
32	0.33628	0.25038	0.18808	0.13673
34	0.32376	0.24164	0.18176	0.13223
36	0.31254	0.23376	0.17604	0.12816
38	0.30240	0.22661	0.17082	0.12443
40	0.29320	0.22008	0.16604	0.12101
42	0.28478	0.21408	0.16165	0.11786
44	0.27705	0.20855	0.15758	0.11495
46	0.26992	0.20342	0.15381	0.11224
48	0.26332	0.19866	0.15029	0.10971
50	0.25718	0.19422	0.14701	0.10734
55	0.24353	0.18430	0.13965	0.10204
60	0.23186	0.17576	0.13330	0.09746
65	0.22172	0.16830	0.12775	0.09344
70	0.21281	0.16173	0.12283	0.08988
75	0.20490	0.15586	0.11844	0.08670
80	0.19781	0.15060	0.11449	0.08383
85	0.19141	0.14583	0.11092	0.08123
90	0.18559	0.14149	0.10765	0.07886
95	0.18028	0.13752	0.10466	0.07668
100	0.17539	0.13386	0.10191	0.07468
110	0.16671	0.12734	0.09699	0.07110
120	0.15920	0.12169	0.09273	0.06799
140	0.14681	0.11234	0.08565	0.06283
160	0.13692	0.10485	0.07998	0.05868
200	0.12196	0.09350	0.07137	