

文長分布の対数正規分布性に関する一考察

——芥川と太宰を事例として——

新井 皓 士

一 文の長さとは対数正規分布

本稿は文長分布の正規性に関する探索的研究の経過報告である。きっかけは「青空文庫」⁽¹⁾に収められた芥川龍之介の一作品に偶然出会ったことであつた。

文の長さ(以下、文長と略す)の分布は一般に、観測値(正の整数)をそのままヒストグラムに描くと極端に左に歪んだ形をとるが、対数変換するとはほぼ正規分布の形をとることを、筆者は「ゲーテ・シラー往復書簡集の文体論的特性」などで示した。⁽²⁾これは語の数で文の長さを測定する欧文、特にドイツ文に限られることではなく、体系的に調べたわけではないが、他の作者や他の言語資料についてもある程度検証した上でのテーゼであつた。⁽³⁾

ただ文の長さは一見して明らかなが多く(たとえばカントあるいは判決文には極端な長文が多いとか、志賀と谷崎の違いとか)、あえて精密な分析をおこなっても作者判別などにさほど強力な手段になりえない可能性が強いと思われ、埋め語(統語機能語)を中心とする多変量解析に主たる関心があつた当時の筆者は、対数正規分布の指摘と分散分析の一部応用程度に留めて、文長分布の問題には深入りを避けていた。しかし講義(言語情報解析論)との関連で入手しやすいデータとしてとりあげた芥川龍之介の遺作『ある阿呆の一生』が、文長分布に関する筆者のやや固定観念化した安直な認識を震撼させたのである。

図 1.3

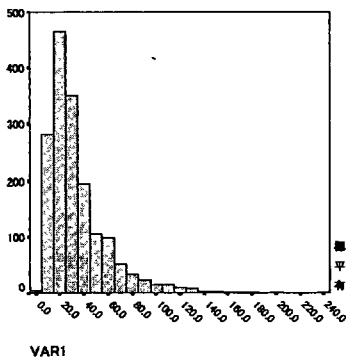


図 1.1

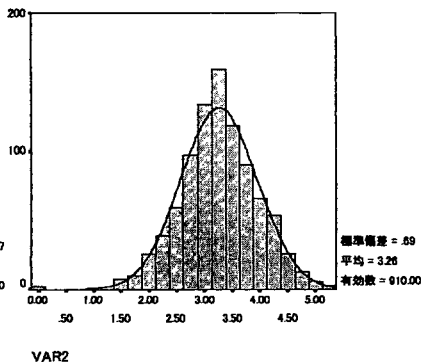


図 1.4

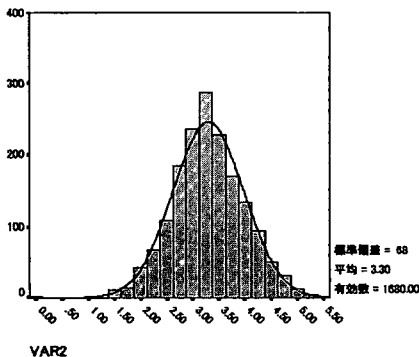
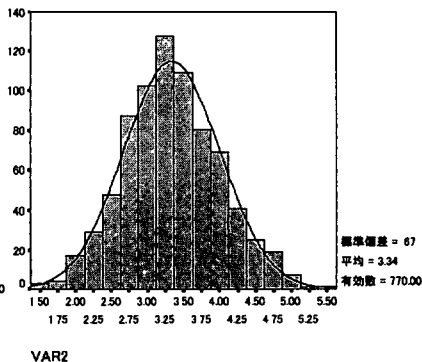


図 1.2



図一・一、図一・二は文長分布が対数正規分布に従うことを示す典型的な例で、『徒然草』(吉田兼好⁽⁴⁾)の上巻と下巻から章段を分ける文言を除いた漢字かな混じり文をデータとし、句読点を省き文字数のみで文長を測っている。現代日本語文では感嘆符や疑問符を含め様々な記号が用いられるが、これらはいわば補足的な視覚情報として作者の意思を伝えようとするもので、文の解釈の方向付けに寄与するとはいえず、文字言語のぎりぎりの使命である音声転写という意味からは省略可能である。無論「ふたえにまげてくびにかけるじゆず」の故事もあるから、なお検討が必要かもしれないが、少なくとも古典との連続性を重視する限り、句読点(これらは洋の東西

をとわず比較的新しいものである)等はさしあたり数え上げない方針をとるべきかと筆者は考えている。もっとも平安時代のかたな文学に関しては原則としてひらがな(音節)単位で扱うことを方針としている立場からすれば、漢字かな混じりについても厳密にはあらためて検討すべきところであるが、ここでは漢字表記部分を強いてかなにひらくことをせず、入手した資料のままに受け取り、もって漢文訓読調と和文調を混成させた鎌倉時代の代表的文人の筆致を尊重することとする。なお、図一・三は『徒然草』上下を合わせた全体について観測値(正の整数)をそのままヒストグラムで表したものの、図一・四はその自然対数をヒストグラムにしたものである。

対照してみれば実測値(図一・三)と対数変換値(図一・四)の示す分布の型が異なることは明らかである。後にくれるように、大量のデータに関する正規性の検定としては相当厳しいリリフォア検定(リリフォアの修正に基づくコルモゴロフ・スミルノフ検定)でも、対数変換したデータによる徒然草上(図一・一)、徒然草下(図一・二)、徒然草全巻(図一・四)のいずれも、観測値の数(標本の大きさ)が増すにつれ有意確率こそ低下す

るとはいえ、正規分布に従うことが認められる。これによって正規分布を前提とする統計的検定への道もひらけることになるが、たとえば徒然草の、上巻と下巻の文長の平均は等しい、という帰無仮説を立て、 t 検定を行なってみると、両者の分散が等しい場合であれ等しくない場合であれ(分散に関しては本来レーベンの検定を行なうべきだが、ここでは省略)、求められた検定統計量は t 分布の一点より小さく、仮説は採択(保留)される。直観的にも両者の平均値はほぼ等しいのだが、この検定によって直観的判断に客観性が認められることになる。

本文に入る前に、日本語の文を文字数で測定することに関連して、『古事記』と『源氏物語』の手元のデータの分析から気づいたことを、簡単に報告しておこう。『源氏物語』については刊行書に基く漢字かな混じりの全文データもあるが、先にふれた原則に従ってすべてを平仮名に直す時間的余裕が目下ないので、数年前にいわば層別系統抽出標本の形で用意した自前の本文データを使い、⁽³⁾当時は気づかなかった対数変換による文長分布を調べてみると、いずれも対数正規分布に従う、という結

果を得た。すなわち平仮名文の長さの分布も対数をとると正規性を示す一例を得たことになるが、漢字を全文に用いて日本語(ヤマト語というべきかもしれない)を表記した『古事記』の場合はどうであろうか。これは選録者である太安万侶も序文に述べているように、全体としては漢文方式の文を骨格としながら、漢字を音仮名、訓仮名、および慣習表記の三様に使い分けて事物や人の名称、思念などを表記する苦心の試みであり、一般に流布する書き下し文の元である原文は、複雑で多元的な原理に従った漢字表記文である。⁽⁶⁾それゆえ事前の予測では到底正規分布には遠いと思われたが、少なくとも『古事記』中巻(便宜上後世の名称でいえば神武より応神天皇までを含む)は、リリフォー検定では対数正規性が否定されるが、カイ二乗適合度検定では正規性の仮説が棄却されないという、予想外の興味ある結果を得た。また試みに移動平均(二十項)をとってみると、上巻、中巻ともに、その文長分布は対数正規分布に従うことも明らかになった。時間的制約もあって下巻の分析には未だ至っていないが、元のデータでは正規性が否定される上巻について、『古事記』成立期のヤマト朝廷に直結するア

マテラス系「神話」とは些か趣きを異にする出雲神話(スサノヲ・オホクニヌシ系)の部分を除いてみると、残余の文長分布は対数正規分布に従うことが認められる。これは内容の上からも異質ないし古層の伝承性が感じられる出雲系神話の特殊性が文長分布に或いは反映しているかとも考えられるが、ともあれ『古事記』分析を通じて筆者は、大きな標本の場合、移動平均(二十項)をみる効用に気づき、またコルモゴロフ・スミルノフのノンパラメトリック検定とリリフォー修正、及びカイ二乗適合度検定を比較検討する必要があることとなった。詳細は後段に譲ることとして、ここではかな表記、漢字表記の日本語古文でも文長の対数正規分布が認められうること、特に移動平均をとった場合は、その可能性が増すこと、また短いながら全文が正統的漢文で書かれている太安万侶の序文も対数正規分布に従っていることを指摘するにとどめる。

二 芥川晩年の作品と文長分布

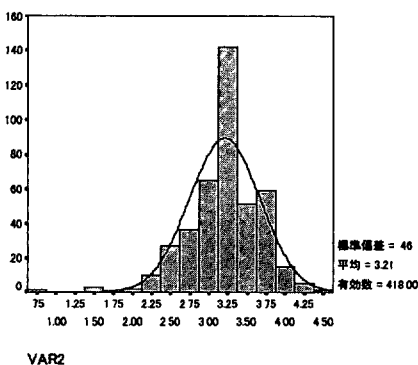
芥川龍之介は一九二七(昭和二)年七月、東京田端の自宅において致死量の睡眠薬を仰ぎ三十五年の生涯を自

ら閉じた。妻や友人菊池寛などにあてた遺書の他に、友人久米正雄に処置を一任する原稿『或阿呆の一生』を残していた。五一の断章から成り、「彼」が成人以後の断片的思いつきを書き記した形をとっている。久米宛のことづて（緒言）に、「僕は今最も不幸な幸福の中に暮している。…（中略）…ではさようなら。僕はこの原稿の中では少なくとも意識的には自己弁護をしなかったつもりだ。…（中略）…どうかこの原稿の中に僕の阿呆さ加減を笑ってくれ給え。」とある。「彼の前にあるものは唯癡狂か自殺かだけだった」という心境だったらしい。良識ある「都会人」でありつつ作家として真摯に創造的使命を担う十年余りの生活に心身とも疲労困憊していたかに思われる。しかしこれらの事情は文学青年的常識であるうし且つ真相は幾重にも秘密のヴェールに覆われているであろうから、これ以上一知半解の言及をする必要はあるまい。ここでふれようとしたのは筆者（新井）が意図的にこの作品を取上げたのではなく、偶然の再会にすぎないということであり、偶々その文長分布のヒストグラムを描いてみて、その特異な形に驚き、ついずると他の芥川作品に手を伸ばし、ひいては文長の対数正規分

布の検証という深みにはまったということである。偶然の再会、と書いたが、むろん筆者も知識人のはしくれて芥川の作品は中学以来多少とも目を通していた。しかし作品に漲る緊張感に疲れるので、若いうちも漱石や井伏ほど親しみは感じられなかったし、ましてや現実なるものに日々直面する中年以降においてをや、という具合、今回も講義の題材に使うつもりで「青空文庫」をのぞき、五十音順で並ぶ作家の中から、留学生の受講者もいることから、まずは無難な選択のつもりで芥川をつい選び、続いてまた五十音順に並ぶ作品の中から早めに出る『秋』と『或阿呆の一生』を選んだのである。軽率な選択ではあった。『或阿呆の一生』は内容的には些か陰鬱の感があり、格助詞「に」の使い方がやや特殊であり、必要以上に「或る…」という表現が頻出するし、到底「笑う」わけにはいかない。『秋』は芥川独特の気迫が薄く、なにか精神的弛緩のようなものを感じる。つまり統計分析の資料とはいえず、教材としては大正期を代表する芥川のもっと優れた端然とした作品を選ぶべきではあった。

図二は『或阿呆の一生』の本文（小見出しを除く）文

図 2



長分布(対数)のヒストグラムである。文長分布に限っていえば、このように中央部が異常に突出した分布を筆者はこれまで見たことがなかった。僅かにカフカの『変身』(Verwandlung)が中央部がかなり突出する例として思い出されるが、これほど極端ではなかった。筆者は一種の不安と予感を感じながら、「青空文庫」に収められている芥川作品の中で一定以上の長さをもつものを次々とダウンロードし、エディタやAWKを用いて分析用のテキストデータとして整備し、最終的には百八編の

作品について、文長を文字数で計測し対数変換したデータによる、各作品の平均値、中央値、標準偏差などを調べてみた。標本数がこれほど増えてくると歪度や尖度も含めAWKの手製スクリプトでカヴァーするのは億劫となり、途中から統計的基礎データをとるのにもSPSSを用い、また一標本ノンパラメトリック検定(正規性のコルゴモロフ・スミルノフ検定)を行なって一覧表に記録した。表一はその一部で、芥川が創作活動を始めた一九一五(大正四)年から五年間に発表したものと、晩年の二年間に当る一九二五(大正一四)年以後のものを示し、間の年月の作品は(空白行)は省いてある。第一行目にあるnは当該作品の文の数(標本の大きさ)、次の二列は対数変換したデータによる平均値と標準偏差で、以下これに基いて算出されている。最後の二列がコルモゴロフ・スミルノフの正規分布検定で、検定統計量(Z)と漸近有意確率(asymptotic significance)を示している。但しこれはリリフォアの有意確率修正を考慮していないので、「保守的」(甘い、と言うべきかもしれない。conservative)とされる判定結果である。

表一の第九列を順に見ていくとすぐに気づくことは、

(7) 文長分布の対数正規分布性に関する一考察

表1 芥川作品の文長分布データ (一部)

タイトル	n	mean	標準偏差	変異係数	歪度	尖度	K-S (z)	確率 ^a	sig
羅生門	153	3.361	0.563	0.168	-1.134	3.417	1.09	0.186	
鼻	159	3.373	0.539	0.160	-0.789	1.509	1.007	0.263	
酒虫	186	3.324	0.741	0.223	-0.666	0.542	1.391	0.042	
仙人1	136	3.467	0.655	0.189	-0.265	0.431	0.941	0.338	
芋粥	399	3.259	0.668	0.205	-0.488	0.114	1.246	0.090	
手巾	189	3.354	0.636	0.190	-0.717	0.997	1.317	0.062	
煙草	167	3.300	0.617	0.187	-0.723	0.374	0.923	0.361	
運	203	3.171	0.754	0.238	-0.970	1.401	1.25	0.088	
偷盗	1342	3.358	0.795	0.237	-0.483	0.065	1.981	0.001	
猶太人	143	3.72	0.643	0.173	-0.32	0.03	0.553	0.92	
忠義	391	3.36	0.661	0.197	-0.67	0.756	1.3	0.068	
二手紙	285	3.001	0.501	0.167	-0.522	1.449	1.028	0.242	
大石	206	3.4927	0.619	0.177	-1.277	3.069	1.496	0.023	
戯作三昧	632	3.169	0.632	0.199	-0.97	0.194	1.545	0.017	
隆盛	306	3.28	0.664	0.202	-0.775	0.808	1.503	0.022	
地獄変	467	3.743	0.690	0.184	-0.958	0.896	2.176	0	
開化の夫	278	3.963	0.688	0.174	-0.554	0.201	1.092	0.184	
蜘蛛の糸	59	3.616	0.705	0.195	-1.243	0.613	1.421	0.035	
奉教人	150	3.774	0.666	0.176	-0.927	1.406	1.265	0.082	
枯野抄	122	3.925	0.587	0.150	-1.75	6.783	1.14	0.149	
邪宗門	629	3.878	0.666	0.172	-0.8	0.912	1.985	0.001	
毛利先生	241	3.736	0.755	0.202	-0.745	0.299	1.522	0.019	
あの頃の ギリシボ	358	3.524	0.571	0.162	-0.799	1.595	1.313	0.064	
竜	196	3.84	0.682	0.178	-0.841	0.64	1.39	0.042	
竜	152	3.994	0.755	0.189	-0.557	-0.481	1.314	0.063	
路上	1187	3.379	0.814	0.241	-0.526	-0.109	2.689	0	
疑惑	242	3.769	0.666	0.177	-0.248	-0.245	0.901	0.391	
妖婆	677	3.716	0.837	0.225	-0.77	0.362	2.06	0	
鼠小僧	263	3.525	0.907	0.257	-0.513	-0.242	1.33	0.058	
大導寺	410	3.229	0.497	0.154	-0.902	3.013	2.962	0	
馬の脚	395	3.272	0.643	0.197	-0.721	1.435	2.146	0	
温泉	155	3.595	0.651	0.181	-0.568	0.612	0.807	0.532	
海の辺	218	3.210	0.685	0.213	-0.659	0.302	1.365	0.048	
湖南の扇	292	3.131	0.690	0.220	-1.025	1.162	2.560	0	
カルメン	57	3.297	0.578	0.175	-0.663	-0.201	0.925	0.359	
点鬼簿	134	3.301	0.469	0.142	-1.113	3.027	1.453	0.029	
夢	207	3.334	0.563	0.169	-1.584	4.720	2.001	0.001	
彼	186	3.205	0.621	0.194	-0.795	1.026	1.283	0.074	
彼第2	206	3.007	0.574	0.191	-0.901	1.121	1.547	0.017	
玄鶴山房	308	3.268	0.528	0.162	-1.116	2.858	2.519	0	
蜃気楼	192	2.990	0.658	0.220	-0.945	0.622	2.005	0.001	
河童	1166	3.297	0.648	0.197	-0.906	1.632	3.586	0	
誘惑	315	3.061	0.604	0.197	-0.508	-0.484	2.361	0	
浅草公園	326	3.016	0.552	0.183	-0.360	1.243	1.254	0.086	
文芸的な たね子の	1414	3.468	0.417	0.120	-0.171	0.646	2.861	0	
冬	119	3.154	0.635	0.201	-1.285	2.379	1.771	0.004	
手紙2	176	3.391	0.596	0.176	-1.490	4.242	1.565	0.015	
手紙2	177	3.324	0.545	0.164	-0.730	1.475	1.235	0.095	
西方の人	450	3.313	0.404	0.122	-1.118	4.980	2.943	0	
蘭中間答	310	2.743	0.497	0.181	-0.502	2.366	1.972	0.001	
続西方	241	3.450	0.381	0.110	-0.356	1.069	1.548	0.017	
幽車	835	3.195	0.515	0.161	-0.998	1.865	4.322	0	
或阿呆の	430	3.200	0.464	0.145	-0.882	0.235	2.655	0	
侏儒の	1108	3.330	0.499	0.150	-0.549	1.161	2.829	0	

表2 芥川作品における文長分布の正規性
(K-S検定における有意確率0.000の頻度)

年次別分布状況

年度	作品数	0—作品	比率
1916以前	10	0	0
1917	7	0	0
1918	7	1	0.14
1919	7	2	0.29
1920	13	4	0.31
1921	6	2	0.33
1922	14	2	0.14
1923	8	3	0.38
1924	9	1	0.11
1925	4	2	0.50
1926	4	1	0.25
1927	16	7	0.44
その他	3	1	0.33
計	108	26	0.24

標本の大きさと漸近確率

文数	as>=0.000	as=0.000
n<=100	3	0
n<=200	33	0
n<=300	31	2
n<=400	8	11
n<=500	2	4
n<=1000	4	5
n>1000	1	4
計	82	26

漱石の賞賛をうけた『鼻』をひっさげて颯爽とデビューした当初は漸近有意確率〇、つまり正規性の仮説が〇・一％水準で棄却されるのがみられない(最初の例である『地獄変』は大正七年)のに対し、晩年の二年間にはそれが頻発することである。また『或阿呆の一生』と同様に中央部が極端に突出するヒストグラムとなる作品、『大導寺信輔の半生』、『玄鶴山房』、『西方の人』、『齒車』はすべてこの晩年に集中している。その先頭にあり自伝的要素を含むとされる『大導寺信輔の半生』は関東大震災⁽⁹⁾の翌年暮れに書かれたもので、一九二五(大正十四)年一月に発表されるが、すでに数年来健康が悪化しつつあったことに加えて死の前年であるこの年は『近代日本文芸読本』編集問題のこじれから大きな精神的打撃を受け、更に翌年早々に義兄の不幸(火事と鉄道自殺)に見舞われて後始末に奔走し、遂には「前にあるものは唯発狂か自殺」という状況に追いこまれたものと憶測される。芥川作品の文長分布の変化と彼の心身の消耗、そして自殺との間には何らかの関連がありそうに思われるが、結論を急ぐ前に我々はもう一度データを見直し、正規性の確率が

零に近くなる要因は他にみられないか検討してみることにする。

表二の左側は、文長分布の正規性について調べた作品数と非となる作品数を年次別に示したもので、大体において晩年（と言っても享年三五の一生だが）に近い年の方が正規分布とみなしうる作品の比率が低くなっている。但し一九二六（大正一五・昭和一）年が例外的であるのは、死の前年にあたるこの年は発表された作品が少ないばかりでなく、比較的短いものばかり（全て文数にして三百以下）であり、特に『カルメン』などは文数が六〇にも満たない。このような場合は右側の表に明らかかなように、正規分布の仮説が棄却されることはほとんどない、と言ってよい。また「その他」とあるのは『侏儒の言葉』のように数年にわたるものや小編の寄せ集めである。K-S検定の結果が文の数（標本の大きさ）によって異なる傾向を示すことは右表から明らかで、文数が四百を越えると分布の正規性が認められないケースが多くなり、千を越えるとはほとんどが正規分布の仮説棄却となる。これは n の大きさが一定限度を越えると、K-S検定の有意点は近似値の式で与えられ、 n が大きくなるにつれ厳

しい条件となるからであろうか。「保守的」なコルモゴロフ・スミルノフでそうであるから、リリフォアの修正値による場合は更に厳しく、初期を別とすれば一定以上の長さをもつ芥川作品の相当数は文長が正規分布に従うとは言えなくなる。これはヒストグラムから得る直観的印象と——標本が小さいとヒストグラムはかなりばれた形でも検定結果は正規性の仮説採択となり、標本が大きいとかなり整った釣鐘型を示しているのに検定結果は仮説棄却となりうるという意味で——やや矛盾する。この問題については第四節においてあらためて考察することとし、とりあげた芥川作品百八のうちK-S検定で正規性を認められぬものが二六、そのうち十一が晩年の三年足らずの間に集中していること、またヒストグラムにおいて極めて特殊な形をとるものがこの期間に集中して現われることを確認して、次節に移ることにしよう。

三 太宰作品の文長分布と対数正規性

芥川における文長分布の正規性の乱れとでも呼ぶべきものが、自殺に至る精神風景と何らかの関係を有するものか、という疑問が、正規分布の検定に関する方法論的

疑問と並んで、筆者の心をとらえ始めた以上、類似したケースとの比較対照を怠ることはできない。となると創作に対する真剣さと力量において匹敵する自殺作家としてまず考えられるのは太宰治である。筆者は早速また「青空文庫」のお世話になり、せっせとダウンロードを始めた。

もっともこの両者は一見かなり違っている。東京育ちの龍之介はしがない官吏生活をする母方の養子となつて芥川を名乗りそのまま筆名にも使つたが、多額納税者として貴族院議員になる父をもつ津軽の旧家の六男坊、津島修治は、地主階級を恥じて非合法運動に首をつっこんだり投身心中の生き残りとして恥をさらした後に、太宰治のペンネームで創作活動に専念する。実名と筆名は小市民社会にそれなりに付き合う律儀さと文士風無頼生活の放埒さの違いとなり、酒や女に対する態度にもあらわれるといえるかもしれない。歳の差が十七年ある両者の直接的接点はないが、旧制高等学校生徒であつた津島修治は芥川の自殺の報に大きな衝撃をうけたといふ⁽¹⁾。

とにかく太宰は三度自殺を企て、二度心中を試み、二度目の時に事は成就したが、妻子があり、子をなした愛

人もあるのに、心中相手は知り合つてさほど間もない間柄だつた。胸部疾患、咯血、不眠症、戦後混乱期とはいへ、ひどい話で、筆者は太宰を罵倒したという三島も好きではないが、軟弱無頼は真の文学を貶めるものとして、健全な市民精神に満ち満ちた一橋では紹介するのも気がひけるくらいに感じていた。しかしどこかに魅力があるのだろう。『桜桃』という太宰晩年の作品などを読むと、側側と胸をうたれるような気がすると同時に手前勝手もいい加減にしろ、と俗人は言いたくなるが、没後五十年経つ最近まで「桜桃忌」なるものが続き忍び音にホトトギスも鳴いて血を吐いたともいふ。

さて芥川にくらべて太宰については資料分析が未だ充分ではないが、表三はその暫定報告である。一見して明らかのように、太宰の場合、第十一列のK-S検定による漸近有意確率が〇となっているものは、一例しかみられない。芥川の場合百八分の二六、二四%の〇確率作品があつたが、こちらは約三六分の一、三%弱の率であり、太宰の方は資料がなお不足している——たとえば代表作の一つ『斜陽』が欠けている——ことを考慮しても、歴然とした差があることは否定できない。すなわち、無頼

(11) 文長分布の対数正規分布性に関する一考察

表3 太宰作品の文長分布データ (一部)

発表年	題	n	m	SD	CV	Min-Max	K-S/Lil' z	Lilief signi	K-S z	K-S as
1933(S08)	列車	63	3 5996	0 6781	0 188	2 08-5 19	0 069	0 200	0 547	0 926
1934(S09)	彼は昔の	1161	2 7714	0 8525	0 308	0 00-5 12	0 050	0	1 691	0 007
*	ロマネスコ	590	3 1213	0 7570	0 243	1 10-5 33	0 036	0 062	0 881	0 420
1935(S10)	逆行	453	2 8912	0 7398	0 256	0 00-5 21	0 045	0 229	0 956	0 320
*	道化の華	1448	2 7535	0 5486	0 199	1 39-3 83	0 054	0	2 049	0
1936(S11)	狂言の神	414	3 0898	0 9653	0 312	0 69-5 22	0 033	0 200	0 681	0 743
1937(S12)	Human L	443	2 7525	0 8415	0 306	0 00-5 29	0 048	0 017	1 006	0 264
1939(S14)	富嶽百景	439	2 9534	0 9165	0 310	0 69-5 32	0 050	0 011	1 042	0 228
*	皮膚と心	368	3 1451	1 0131	0 322	0 00-5 66	0 034	0 200	0 648	0 796
1940(S14)	兄たち	143	3 6695	0 8325	0 227	1 39-5 60	0 069	0 091	0 827	0 502
*	駆込み訴	450	2 9874	0 8629	0 289	0 69-5 55	0 083	0	1 757	0 004
*	善魔を	476	2 9405	0 8011	0 272	1 11-5 24	0 048	0 010	1 050	0 220
*	乞食学生	1396	2 7411	0 7455	0 272	0 69-5 70	0 038	0	1 437	0 032
1941(S16)	清貧譚	331	2 9625	0 7209	0 243	1 10-5 03	0 048	0 066	0 869	0 437
*	千代女	199	3 6041	0 8766	0 250	1 79-6 34	0 059	0 094	0 826	0 502
1942(S17)	恥	199	3 6041	0 8766	0 250	1 79-6 34	0 059	0 094	0 826	0 502
*	花火	638	2 6738	0 7770	0 291	0 69-5 35	0 042	0 009	1 064	0 208
1943(S18)	黄村先生	395	2 8501	1 0401	0 365	0 69-6 09	0 069	0	1 370	0 047
*	実朝	916	3 8561	1 2421	0 322	0 69-6 98	0 048	0	1 448	0 030
1944(S19)	散華	299	3 1512	0 7610	0 242	0 69-4 29	0 058	0 018	0 995	0 275
*	津軽	2902	3 0996	0 8742	0 282	0 69-5 93	0 024	0	1 313	0 064
1945(S20)	竹清	244	3 0174	0 8603	0 285	0 69-5 18	0 043	0 200	0 670	0 760
*	惜別	2005	3 2217	0 9517	0 295	0 69-6 30	0 025	0 005	1 122	0 162
*	お伽草紙	2066	3 0681	0 8443	0 275	0 00-6 04	0 035	0	1 593	0 012
1946(S21)	未帰還の	339	2 9805	0 8576	0 288	0 00-5 64	0 049	0 050	0 394	0 049
*	親友交歓	486	2 9606	0 8392	0 284	0 69-5 83	0 036	0 168	0 796	0 550
1947(S22)	トカント	247	3 3539	0 9917	0 296	0 69-6 20	0 061	0 200	0 577	0 893
*	母	271	2 7808	0 8202	0 295	0 69-5 27	0 059	0 022	0 976	0 296
*	ヴィヨンの	374	3 2578	1 0968	0 337	0 00-6 53	0 040	0 194	0 783	0 573
*	父	233	2 9290	0 8992	0 307	0 00-4 75	0 047	0 200	0 722	0 675
*	おさん	187	3 3205	1 0514	0 317	0 69-5 90	0 077	0 008	1 058	0 213
1948(S23)	女類	223	2 9833	0 9258	0 310	1 10-5 47	0 064	0 025	0 962	0 312
*	桜桃	145	3 0072	0 8721	0 290	0 69-5 12	0 061	0 200	0 732	0 657
*	グッドバイ	586	2 7233	0 8293	0 305	0 69-4 94	0 05	0 001	1 214	0 105
*	家庭の	230	3 0424	0 8786	0 289	0 69-5 85	0 064	0 022	0 976	0 297
*	人間失格	1459	3 2308	1 0855	0 336	0 00-6 26	0 021	0 135	0 788	0 564

派太宰の文長分布は高度に対数正規分布に従うのである。と同時に変動係数が最低〇・一八八、最高〇・三六五と、芥川(最低〇・一一、最高〇・二五七)にくらべてかなり高いことは、太宰の文の方が、ある意味では律動的、ある意味では落ち着きがないことを示しているといえようか。いずれにしても太宰の晩年作品は、文数が千五百に近い『人間失格』や未完の遺作ともいうべき文数六百に近い『グッドバイ』を含めて、その文長分布が対数正規分布に従う、ということができよう。変動係数とあわせて考えてみると、単調で堅苦しいところがなく、律動感がありながら全体としてはバランスがとれている、といえようか。意外な結果である。死を前にした極度の緊張のようなものが文体からは感じられないのである。あるいは度重なる自殺・心中未遂を経て微妙な安定の域に達し、「滅びのアカルサ」(『実朝』)のようなものが反映して

表4 分布の正規性に関する Kolmogorov-Smirnov/Liliefors の検定
(データ: 芥川「カルメン」)

文長[対数度数]	累積度数	累積比	z-score	Norms(z)	Fi-D(i-1)	Fi - Di
1.79176	1	1	0.01754	-2.6038	0.0046	0.004609
1.94591	1	2	0.03509	-2.3371	0.0097	0.007827
2.07944	1	3	0.05263	-2.1061	0.0176	0.017488
2.30259	1	4	0.07018	-1.7200	0.0427	0.009912
2.56495	2	6	0.10526	-1.2660	0.1028	0.032581
2.63906	3	9	0.15789	-1.1378	0.1276	0.022344
2.70805	4	13	0.22807	-1.0184	0.1542	0.003650
2.77259	1	14	0.24561	-0.9067	0.1823	0.045794
2.83321	1	15	0.26316	-0.8018	0.2113	0.034288
2.89037	2	17	0.29825	-0.7029	0.2411	0.022108
2.94444	1	18	0.31579	-0.6094	0.2711	0.027109
2.99573	1	19	0.33333	-0.5206	0.3013	0.014476
3.21888	1	20	0.35088	-0.1345	0.4465	0.113160
3.25810	2	22	0.38596	-0.0667	0.4734	0.122548
3.29584	2	24	0.42105	-0.0014	0.4995	0.113492
3.33221	2	26	0.45614	0.0616	0.5245	0.103493
3.40120	3	29	0.50877	0.1809	0.5718	0.115652
3.43399	1	30	0.52632	0.2377	0.5939	0.085162
3.46574	1	31	0.54386	0.2926	0.6151	0.088774
3.49651	1	32	0.56140	0.3459	0.6353	0.091414
3.52636	3	35	0.61404	0.3975	0.6545	0.093100
3.58352	1	36	0.63158	0.4964	0.6902	0.076162
3.61092	1	37	0.64912	0.5438	0.7067	0.075137
3.63759	1	38	0.66667	0.5900	0.7224	0.073268
3.68888	2	40	0.70175	0.6787	0.7513	0.084673
3.71357	2	42	0.73684	0.7214	0.7647	0.062925
3.73767	1	43	0.75439	0.7631	0.7773	0.040466
3.78419	2	45	0.78947	0.8436	0.8006	0.046174
3.80666	2	47	0.82456	0.8825	0.8112	0.021775
3.82864	1	48	0.84211	0.9205	0.8214	0.003207
3.85015	2	50	0.87719	0.9578	0.8309	0.011200
3.87120	2	52	0.91228	0.9942	0.8399	0.037261
4.06044	1	53	0.92982	1.3216	0.9069	0.005428
4.11087	2	55	0.96491	1.4089	0.9206	0.009260
4.14314	1	56	0.98246	1.4647	0.9285	0.036414
4.15888	1	57	1	1.4919	0.9321	0.050313
MAX[G,H]						
n=57	m=3.296624					statistic D
	sd=0.577942					0.122548

Liliefors' test table of critical value (n over 30)

$\alpha=0.05$	$0.89/\sqrt{57}=0.11788$	$dn < D$	reject
$\alpha=0.01$	$1.04/\sqrt{57}=0.13775$	$dn > D$	accept

Kolmogorov-Smirnov' test table of critical value (n=57)

$\alpha=0.05$	$dn= 0.15906$	$dn > D$	accept
$\alpha=0.01$	$dn= 0.19758$	$dn > D$	accept

いるのだろうか、師とする井伏鱒二の自ずからなる蕉陶のお蔭であろうか。もっとも太宰作品の中でK-S検定で漸近有意確率が〇となる唯一の作品『道化の華』は、自殺未遂、心中の死に損ない、自殺未遂と続いた後の二六歳の作品ではある。表二でみたように、芥川の場合は標本の大きさ(文数)が千を越える作品五点のうち四点がK-S検定で漸近有意確率〇となったが、太宰では同じく千を越える作品六点のうち僅か一点だけということになる。これは芥川の場合が異例であって、文長は対数正規分布に従う、というテーゼは太宰の場合のように、大標本に関しても一般に肯定されるとみてよいのだろうか。そもそも大標本の正規分布検定においてコルモゴロフ・スミルノフの「保守的」有意点をとるべきなのか、それとも厳しいリフォアの修正値により判断すべきか。後者をとれば太宰作品もまた六点中五点が正規性仮説棄却という結果になるが、ヒストグラムなどからくる視覚的印象も軽んじることなく、データを適当にクラスわけしてカイ二乗適合度検定を行なうべきであろうか。これらの問題を最終的に判断するには、まず様々な文のデータを集め、それぞれの方法を適用して総合的な判断をす

べきであろう。次節ではさしあたり手元のデータにより、方法的検討を試みることにする。ともあれ文長分布の正規性でみる限り、芥川の場合は初期ないし青年期の調和体が次第に崩れ晩年に至って頂点に達するが、その転回点をなすのは、結婚し大阪毎日新聞社社友となり作家生活が一見安定するかにみえた頃の作、いわば芸術至上主義の画師良秀の不吉な業を描く『地獄変』であるように思われる。一方太宰の場合はむしろ初期に文長分布の非正規性がみられ、晩年には安定した正規分布性を示している。もし自害自決に至る精神状態と文体分布の正規性の間に多少とも関係があるとするならば、太宰は度々暗示している自殺願望を少しづつ手なづけ、文章の世界を安定させていたのかもしれない。しかしこれは輕輕に断言すべき問題ではない。

四 文長分布の検定と移動平均

標本の大きさが三十以上であるとき、その分布が正規分布に従うか否かを調べるには、まずヒストグラムを描いたうえで、帰無仮説の検定をおこなうことになるが、代表的な検定法としてあげられるのは、コルモゴロフ・

図 3.1

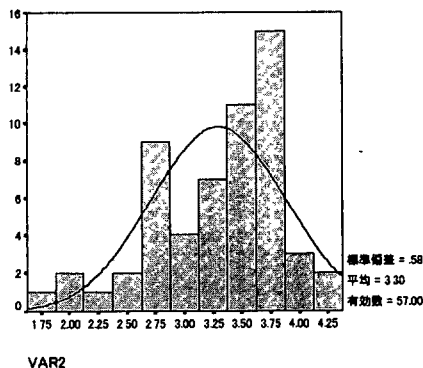
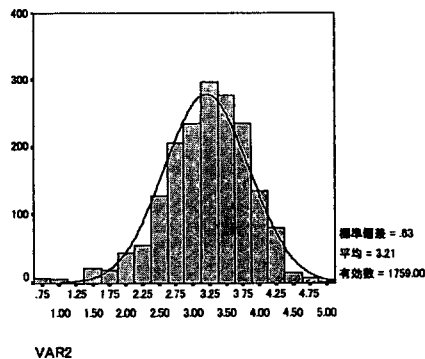


図 3.2



になるほど正規性の仮説は棄却されやすいようにみえるのである。たとえば、次に掲げる図三の一と二ではどちらが正規分布に近いと感じられるだろうか。

常識的には図三・二の方が釣鐘型に近いと感じられるであろう。

しかし検定の結果は、より精確とされるリリフォー検定でも三・一は正規分布に従うと判定され、図三・二はリリフォー修正値による

スミルノフの一標本検定とカイ二乗適合度検定である。いわゆる経験累積分布関数を用いる前者は正規分布のみに応用されるものではないが、正規分布に関してはより精度の高いリリフォーの修正値を用いる検定が最近では優勢であるようで、従来のコルモゴロフ・スミルノフの有意点を用いるのは「保守的」とされることはすでに触れたところである。問題はコルモゴロフ・スミルノフにせよリリフォーにせよ、標本が一定以上の大きさに達すると有意点が通常は近似値の式で与えられる事で、大標本

検定では正規性を否定され、「保守的な」コルモゴロフ・スミルノフの有意点を使う場合、○・一%水準で辛うじて正規性の仮説が採択される。原因は標本の大きさにあるといっただろう。一図は芥川作品『カルメン』で文数は五七、二図は漱石の『心』第一章で文数一七五九。芥川の晩年は文長分布が正規分布から外れる例が多く、漱石は新聞連載の影響もあってか文体が安定しているとみなされるが、ここではまったく逆の様相を呈している。筆者はこのことから、少なくとも文長分布

を検証する場合には、リリフォー修正値によるコルモゴロフ・スミルノフ検定だけで判断するのは危険性を伴うと感じ、大標本（といっても大きさ五百を越えるあたりから）を扱うに際してはカイ二乗適合度検定をも併用すべきかと考えるようになった。両者の基本的な違いはコルモゴロフ・スミルノフが個々のデータをすべて直接使うのに対し、カイ二乗適合度検定はデータを幾つかのクラスに分け度数としてまとめて扱う点にあり、従って計算手数上は少数「自由度 (degree of freedom)」の後者の方が圧倒的に簡便であるはずだが、実際問題としては後者はクラスの設定の仕方が存外厄介で、標本に依じて有効な設定⁽¹³⁾をする必要があり、出来合いのアプリケーションにそっくり頼ることも大概不可能である。苦心の末うまくいって疑問が氷解することもあるが、たとえば上の『心』の例では微妙な差で予想ははずれ、何となく腑に落ちないままにクラス設定を色々変えて試してみたが、その都度結果は正規性の仮説が棄却されるものであった。筆者のこれまでの実験では適合度検定の結果は、コルモゴロフ・スミルノフの未修正値による判定とリリフォー修正値による判定の間に落ちることが多いようだ

が、正規分布仮説の棄却か採択（保留）かではリリフォー修正値による判断との一致度がより高いようである。

コルモゴロフ・スミルノフ検定を補足するものとして、カイ二乗適合度検定のほかに筆者が試みたのは、移動平均をとってみることである。これは元来、文の流れを時系列的にみる為にスク립トを書いたものだが、ふと思いついたことがあって、芥川、太宰それぞれの作品の中で、コルモゴロフ・スミルノフ検定で漸近有意確率が〇となるものについて二十項移動平均をとり、その結果をヒストグラムにしたり検定にかけてみたところ、ひとつ予想外の傾向がみられることが明らかになった。すなわち元のデータでコルモゴロフ・スミルノフ検定（未修正値）の漸近有意確率〇となったものが、二十項移動平均をとると有意確率が多少とも高くなる傾向は芥川作品にみられ、太宰の場合は移動平均をとることで有意確率が下がる例が多いことである。太宰の場合元のデータでは『道化の華』が唯一、リリフォーの修正値ばかりかK-S未修正値によっても正規性が認められないものだったが、この作品とユダを描いた『駆け込み訴え』、ヴィヨンと玉川上水がからむ『乞食学生』だけが移動平均によ

って正規性を増し、他(『津軽』『お伽草子』『惜別』など)は有意確率を下げてゐる。芥川方では晩年の作品でも『大道寺』『齒車』『或る阿呆の一生』など晩年の問題作が確率をあげ5%ないし1%水準で正規分布の仮説採択(保留)となる。ただし、これはコルモゴロフ・スミルノフの未修正値による検定の場合で、リリフォー修正値を使う場合は、芥川晩年では『大道寺』だけが辛うじて○・1%水準で正規分布仮説採択(保留)となり、太宰では初期の『道化の華』と『駆け込み訴え』だけが矢張り1%水準で正規分布仮説採択(保留)となる。筆者がこれまで当ってみた限りでは、元のデータがリリフォー検定で正規性の確率が○(近似)と出た場合、二十項移動平均データでも矢張り○と出るのが普通で、これはむしろ例外に属するのだろうか。『大道寺』は芥川晩年の転機をなす作品であり、『道化の華』は太宰が青春と決別する作品といえとすれば、これも奇妙な暗合だ。そう言えば二人とも作家活動の最初に「晩年」やら「老年」やら題する作品を書き、ヴィヨンに惹かれていた。違いは「生活的宦官」と自嘲するか「無頼派」を自認するところからくるものが多少ともあるとすれば、概ね逆

方向をとる両者の文長分布の変化と暗合は何を意味するのであろうか。

たまたま芥川晩年の特異な形をとる文長分布に目を奪われて、正規分布の検定問題に論は波及した。未整理のデータもあるが、紙幅と時間に責めを預けて、ひとまず擱筆することにする。⁽¹³⁾

(1)「青空文庫」は、グーテンベルク・プロジェクトと同様、著作権の切れた文学作品などをデジタル・テキストとして入力し、インターネットを通じて公表、人類の共有文化資産とする趣旨にたつ、自由な人々による自由な知的活動である。芥川作品、太宰作品に関して筆者は今回多大な恩恵をこうむったことをここに記し、心から感謝の意を述べたい。

(2)「文体統計論的にみたゲーテ・シラー往復書簡集の特性」(『一橋論叢』第一二二巻第三号、第一二三巻第三号)、『ラレーブーフ』(二五九七)の文体統計解析(早稻田大学文学研究科紀要第四六輯)。

(3) つとに、C・B・ウィリアムズが文長の分布は大体において対数正規分布していることを示した、とA・ケニイは指摘している(『文章の計量』)。ただ、この書の翻訳はどこどころ意味がとりにくいところがあり、またこの指

摘の少し前では、文長について「正規分布の釣鐘型曲線とはほとんど似通っていない」ともあって、筆者はウィリアムズの名をすっかり忘れていた。安本美典なども対数正規性のことに論及したものがあつたように記憶するが、啓蒙的な前川守『文章を科学する』をはじめユール以来の文長分析が紹介される場合ほとんどが実測値を扱っている印象がある。あるいはこれは自らの課題追及に急で先行研究の網羅的検討に必ずしも充分な時間を割かぬ筆者の近視眼のせいかもしれない。

(4) このデータもまた、インターネットを通じて公表されたコーパスのお蔭をこうむっている。福井大学教育学部の関係者各位に御礼申し上げる。なお後述の『古事記』本文データも同様のルートにより入手し加工利用させていたのだいたものである。

(5) これについては、「源氏物語・宇治十帖の作者問題」(『一橋論叢』第一一七巻第三号)。層別、という意味は全文から全く無差別抽出を行なうと、有力な章が抜ける事態もあるので、各章から必ず採取するようにしたことを仮にこう呼んだのである。

なお、個々の巻では、桐壺、若紫、夕顔、柏木、竹河が平仮名データとして用意してあったので調べてみると、若紫以外はK-S検定(リリフォー修正)で正規分布が認められ、若紫も二十項移動平均をとると正規分布に属すると認められる。

(6) 「文」の区切りは、倉野憲司校注『古事記』(岩波文

庫)に従った。またダウンロードした本文には、原文に割注の形で挿入されている人名の読み方(音仮名)なども入力されているが、これらの注はすべて削除し、加之おそらくダウンロードの際に混入したであろう本文の若干の乱れを正した。また、太安万侶の「序」は上巻(カミツマキ)に併合されているが、本論では別扱いとした。

(7) これも遺作となつた『幽車』にも同様の傾向がみられる。

(8) Gibbons, J. D./Chakraborti, S.: Nonparametric Statistical Inference (3. edition, Revised and Expanded), New York/Basel/Hong Kong, 1992, p. 122

また、Liliefors という名の読み方は、『統計解析ハンドブック(ノン・パラメトリック法)』(田中豊・垂水共之編、東京、一九九九)に従った。

(9) 芥川はほとんど被害を受けなかったが、『或阿呆の一生』三十一に焼け跡の印象を「酸鼻」と書き、「誰もかも死んでしまえば善い」という不気味な言葉を記している。

(10) これは旧制中学の副読本として企画されたものだが、著作権や印税問題がからんで、芥川自身についてもあらぬ噂が流れたらしい。今年(二〇〇〇年)十月五日に、この事に関連して芥川が文壇の長老徳田秋声にあて詫び状を七通も寄せていた事、保管されていたその書簡が徳田家で発見された事が、新聞で報道されている。七通という数は素人には何とも異常に思われる。書く方もさりながら、書かせる方も。よほどの事情があつたのだらうか。翌年には谷

崎との文芸論争もあったようで、『闇中間答』はこのような状況の中で孤立無援の自己模索のあらわれだろうか。

(11) このあたりの伝記的情報はすべて新潮日本文学の年譜や二、三の評伝の受け売りである。

(12) SPSSなどを使うと瞬時に検定結果が表示されるが、補助表を使って検定統計量を算出する例を、表四に示した。エクセルを使用している。

(13) 付表として、芥川・太宰両者の文長平均の正規分布に關するコルモゴロフ・スミルノフ検定を補助表によって行なう例を示した。

(14) 本稿は河村教授の退官記念号に寄稿するものである。従って不吉ともとられかねないことは慎むべきかもしれないが、河村名誉教授の寛仁大度に甘えて以下のことを記す。

本稿の締め切り間際、一步も外出せず徹夜に近い数日が続き脳中朦朧としかけていた時、卒業生からの突然の電話で、社会学研究科教授辻内鏡人氏が急逝されたことを知った。

自転車を利用されている時のことのように、筆者も自転車をよく利用することから氣遣って連絡してくれたのである。そういえば河村先生も車のみならず時々自転車を利用されるようなのでお互い用心して天寿をまっとうしたいもの。

それはそれとして、辻内氏とは構内などで軽く会釈をかわす程度のつきあいではあったが、少し含羞をふくんだような笑顔が彷彿として眼前に浮かんできく。偶々死と関わり深い二人の作家を扱っていたため、殊更印象が深いのであろうか。冥福を祈りたい。(二千年師走)

(一橋大学大学院言語社会研究科教授)

(19) 文長分布の対数正規分布性に関する一考察

(付表)

付表 文長平均値の正規分布に関するコルモゴロフ-Sミルノ検定(補助表)

標本平均	度数	累積度数	累積率	$(A_i - m)/s$	NORMSD	$ F_i - D_i $	ABS(F-D)
2.943	1	1	0.0093	-2.4052	0.0081	0.0081	0.0012
2.732	1	2	0.0185	-2.2602	0.0110	0.0071	0.0075
2.838	1	3	0.0278	-1.6320	0.0513	0.0378	0.0238
2.990	1	4	0.0370	-1.4259	0.0770	0.0492	0.0299
3.001	1	5	0.0463	-1.3822	0.0834	0.0464	0.0372
3.007	1	6	0.0556	-1.3584	0.0877	0.0469	0.0418
3.016	1	7	0.0648	-1.3278	0.0930	0.0374	0.0281
3.019	1	8	0.0741	-1.3109	0.0950	0.0201	0.0209
3.021	1	9	0.0833	-1.3029	0.0963	0.0222	0.0130
3.052	1	10	0.0928	-1.1800	0.1190	0.0357	0.0284
3.061	1	11	0.1019	-1.1443	0.1262	0.0337	0.0244
3.063	1	12	0.1111	-1.0571	0.1452	0.0454	0.0341
3.085	1	13	0.1204	-1.0492	0.1471	0.0329	0.0287
3.095	1	14	0.1298	-1.0095	0.1564	0.0350	0.0287
3.101	1	15	0.1389	-0.9857	0.1621	0.0325	0.0232
3.103	1	16	0.1481	-0.7557	0.2549	0.0960	0.0438
3.130	1	17	0.1574	-0.8868	0.1920	0.0448	0.0336
3.131	1	18	0.1665	-0.7759	0.2190	0.0618	0.0421
3.159	1	19	0.1757	-0.7557	0.2190	0.0490	0.0287
3.161	1	20	0.1848	-0.7478	0.2273	0.0421	0.0328
3.164	1	21	0.1937	-0.7289	0.2309	0.0384	0.0272
3.169	1	22	0.2025	-0.7181	0.2379	0.0353	0.0292
3.171	1	23	0.2112	-0.7082	0.2394	0.0285	0.0172
3.184	1	24	0.2215	-0.6710	0.2588	0.0444	0.0371
3.195	1	25	0.2307	-0.6407	0.2789	0.0285	0.0292
3.201	2	26	0.2398	-0.5892	0.3173	0.0371	0.0188
3.205	1	27	0.2485	-0.5734	0.3302	0.0239	0.0147
3.209	1	28	0.2570	-0.5778	0.2949	0.0180	0.0098
3.210	1	29	0.2653	-0.5535	0.2900	0.0122	0.0079
3.214	1	30	0.2743	-0.5371	0.2954	0.0084	0.0029
3.217	1	31	0.2832	-0.5056	0.3295	0.0327	0.0092
3.221	1	32	0.2919	-0.5099	0.3061	0.0005	0.0006
3.229	1	33	0.3011	-0.4782	0.3183	0.0014	0.0078
3.231	1	34	0.3102	-0.4702	0.3295	0.0070	0.0142
3.234	1	35	0.3192	-0.4584	0.3233	0.0100	0.0192
3.239	1	36	0.3279	-0.4358	0.3305	0.0121	0.0214
3.249	1	37	0.3361	-0.3939	0.3459	0.0089	0.0181
3.251	1	38	0.3442	-0.3910	0.3479	0.0122	0.0225
3.258	2	39	0.3528	-0.3711	0.3552	0.0151	0.0328
3.258	1	40	0.3611	-0.3391	0.3702	0.0307	0.0339
3.259	1	41	0.3692	-0.3532	0.3597	0.0384	0.0477
3.283	1	42	0.3770	-0.3434	0.3657	0.0418	0.0510
3.288	1	43	0.3852	-0.3215	0.3702	0.0465	0.0558
3.298	1	44	0.3932	-0.3238	0.3731	0.0578	0.0670
3.272	1	45	0.4014	-0.3077	0.3792	0.0580	0.0653
3.280	1	46	0.4093	-0.2780	0.3913	0.0532	0.0624
3.297	2	47	0.4172	-0.2088	0.4174	0.0383	0.0548
3.300	2	48	0.4250	-0.1987	0.4220	0.0502	0.0607
3.301	1	49	0.4328	-0.1997	0.4238	0.0671	0.0784
3.313	1	50	0.4405	-0.1451	0.4422	0.0377	0.0670
3.314	1	51	0.4481	-0.1412	0.4439	0.0654	0.0748
3.317	1	52	0.4557	-0.1293	0.4488	0.0699	0.0782
3.322	2	53	0.4632	-0.1055	0.4600	0.0698	0.0683
3.324	2	54	0.4707	-0.1015	0.4596	0.0687	0.1052
3.330	1	55	0.4781	-0.0777	0.4690	0.0958	0.1090
3.332	1	56	0.4853	-0.0658	0.4738	0.1023	0.1096
3.334	1	57	0.4928	-0.0919	0.4753	0.1080	0.1173
3.338	1	58	0.5001	-0.0490	0.4817	0.1108	0.1202
3.354	1	59	0.5073	-0.0174	0.5070	0.0948	0.1042
3.358	2	60	0.5148	0.0333	0.5133	0.0978	0.1183
3.360	1	61	0.5223	0.0412	0.5164	0.1132	0.1224
3.361	1	62	0.5297	0.0452	0.5160	0.1308	0.1301
3.373	1	63	0.5374	0.0939	0.5370	0.1112	0.1204
3.379	1	64	0.5451	0.1168	0.5464	0.1110	0.1203
3.381	1	65	0.5528	0.1042	0.5525	0.1015	0.1127
3.392	1	66	0.5602	0.1081	0.5608	0.1092	0.1184
3.422	1	67	0.5674	0.2871	0.6130	0.0722	0.0815
3.460	2	68	0.5745	0.2381	0.5740	0.0397	0.0582
3.467	1	69	0.5817	0.4655	0.6782	0.0337	0.0430
3.468	1	70	0.5889	0.4895	0.6808	0.0416	0.0508
3.470	1	71	0.5960	0.4744	0.6825	0.0480	0.0573
3.480	1	72	0.6030	0.4508	0.7057	0.0250	0.0443
3.500	1	73	0.6100	0.6201	0.7324	0.0178	0.0283
3.517	1	74	0.6168	0.6638	0.7468	0.0127	0.0219
3.524	1	75	0.6235	0.6315	0.7554	0.0131	0.0274
3.525	1	76	0.6302	0.6955	0.7580	0.0212	0.0304
3.525	1	77	0.6369	0.8144	0.7923	0.0053	0.0040
3.525	1	78	0.6436	0.8200	0.8057	0.0094	0.0061
3.588	1	79	0.6503	0.8860	0.8087	0.0012	0.0001
3.573	1	80	0.6569	0.8858	0.8121	0.0027	0.0116
3.574	1	81	0.6635	0.8886	0.8132	0.0108	0.0201
3.593	1	82	0.6701	0.9651	0.8328	0.0008	0.0008
3.595	1	83	0.6767	0.9730	0.8347	0.0079	0.0119
3.615	1	84	0.6832	1.0563	0.8448	0.0027	0.0065
3.633	1	85	0.6897	1.1237	0.8694	0.0063	0.0009
3.645	1	86	0.6961	1.1713	0.8793	0.0069	0.0024
3.718	1	87	0.7026	1.4528	0.9259	0.0412	0.0200
3.720	1	88	0.7090	1.4687	0.9290	0.0402	0.0309
3.728	1	89	0.7154	1.5321	0.9373	0.0381	0.0288
3.743	1	90	0.7217	1.5593	0.9406	0.0352	0.0229
3.765	1	91	0.7280	1.6471	0.9302	0.0338	0.0243
3.789	1	92	0.7342	1.8530	0.9518	0.0259	0.0187
3.794	1	93	0.7404	1.8626	0.9538	0.0186	0.0085
3.840	1	94	0.7466	1.9445	0.9741	0.0296	0.0204
3.878	1	95	0.7528	1.9630	0.9952	0.0282	0.0190
3.925	1	96	0.7590	2.2815	1.0887	0.0258	0.0185
3.963	1	97	0.7651	2.4327	1.0925	0.0003	0.0010
3.994	1	98	0.7712	2.5551	1.0947	0.0132	0.0040
4.087	1	99	0.7773	1.0000	2.8239	0.9883	0.0075

n=108
m=3.2598
SD=0.2322

$\alpha=0.05$ 0.08584
Likelihood $\alpha=0.01$ 0.10007
K-S(D) $\alpha=0.001$ 0.11737
0.11401

D= 0.13012
D<D accept
D>D reject

表 文長平均値の正規分布に関するコルモゴロフ-Sミルノ検定(補助表)

標本平均	度数	累積度数	累積率	$(A_i - m)/s$	NORMSD	$ F_i - D_i $	ABS(F-D)
2.6738	1	1	0.0278	-1.4063	0.0798	0.0798	0.0502
2.7233	1	2	0.0556	-1.3375	0.108	0.0802	0.0524
2.7411	1	3	0.0833	-1.1937	0.1187	0.0641	0.0335
2.7525	1	4	0.1111	-1.1378	0.1276	0.0443	0.0185
2.7535	1	5	0.1389	-1.1444	0.1283	0.0172	0.0108
2.7714	1	6	0.1667	-0.7034	0.1418	0.0207	0.0203
2.7808	1	7	0.1944	-0.9413	0.1489	0.0178	0.0058
2.8001	1	8	0.2222	-0.8048	0.2105	0.0118	0.0118
2.8012	1	9	0.2500	-0.8948	0.2531	0.0306	0.0303
2.825	1	10	0.2778	-0.8337	0.2581	0.0401	0.0183
2.8405	1	11	0.3056	-0.4984	0.3098	0.032	0.0042
2.8524	1	12	0.3333	-0.4524	0.3259	0.0189	0.0075
2.8626	1	13	0.3611	-0.4278	0.3344	0.0011	0.0027
2.8825	1	14	0.3889	-0.4114	0.3387	0.0244	0.0251
2.8825	1	15	0.4167	-0.3599	0.3884	0.0284	0.0512
2.8833	1	16	0.4444	-0.3504	0.393	0.0526	0.0814
2.8874	1	17	0.4722	-0.3384	0.3883	0.0782	0.1039
3.0072	1	18	0.5	-0.2889	0.394	0.0762	0.108
3.0174	1	19	0.5278	-0.324	0.4075	0.0925	0.1023
3.0424	1	20	0.5556	-0.1488	0.4409	0.0869	0.1147
3.0881	1	21	0.5833	-0.0811	0.4757	0.0789	0.1077
3.0938	1	22	0.6111	0.012	0.5052	0.0782	0.1059
3.1512	1	23	0.6389	0.0484	0.5185	0.0628	0.1204
3.1713	1	24	0.6667	0.1004	0.5479	0.091	0.1187
3.1451	1	25	0.6944	0.2018	0.5799	0.0888	0.1145
3.1512	1	26	0.7222	0.2224	0.588	0.1064	0.1342
3.2717	1	27	0.75	0.463	0.6783	0.0439	0.0717
3.2308	1	28	0.7778	0.494	0.6894	0.0608	0.0884
3.2578	1	29	0.8056	0.5091	0.7111	0.0587	0.0944
3.3205	1	30	0.8333	0.8001	0.7882	0.0174	0.0064
3.3539	1	31	0.8611	0.814	0.8186	0.0127	0.0115
3.5998	1	32	0.8889	1.523	0.8601	0.089	0.0713
3.6041	1	33	0.9167	1.9077	0.8814	0.0726	0.0648
3.6041	1	34	0.9444	1.9877	0.8814	0.0446	0.017
3.6895	1	35	0.9722	1.9808	0.9787	0.0323	0.0045
3.8581	1	36	1	2.8274	0.9957	0.0235	0.0043

n=38
m=3.088
SD=0.2931

$\alpha=0.05$ 0.13479
Likelihood $\alpha=0.01$ 0.17333
K-S(D) $\alpha=0.001$ 0.20333
0.2610
D= 0.1342
D<D accept
D>D reject