

# モンテカルロ法による衆院議席予測精度の検討

鈴木 督久

## 1. 背景と目的

新聞社をはじめとするマスコミ各社は衆議院と参議院の二つの国政選挙において、投票日前に標本調査を実施して選挙情勢の予測報道をする。予測の対象は、政党別の獲得議席数、選挙区別の各候補者の当落、政党別の得票率の三種類である。このうち最も重要な予測が政党別議席数であり、その結果から新聞紙面の大見出しが作られる。

戦後のマスコミ選挙予測報道を再確認すると、国政選挙における議席予測の難しさが示唆される。第一党の「勝利」と「敗北」を大見出しで間違えることもあった。たとえば1979年の第35回総選挙では事前の予測に反して自民党敗北という結果に終わった。近年では1998年の第18回参院選で自民党惨敗の情勢をつかめず予測は大きく外れた。このような状況の中でマスコミ各社は予測失敗の原因を追究し、予測モデルを再構築するために調査・研究を重ねている。

本論文の目的は、選挙予測報道にとって最も重要な政党別獲得議席数の予測精度がどの程度まで達成できるか、という目安を示すことである。これにより、マスコミの選挙予測に対して「当たった」とか「外れた」というような判定をする際にも、客観的な目安と根拠の一つが提供される。

現在のマスコミ各社の選挙予測は、標本調査と統計的データ解析を利用して実施されており、予測精度の評価は主に標本誤差の統計的推測の問題になる。議席予測は政党別獲得議席数の区間予測という形式で表現される。従って、問題は獲得議席数の信頼区間を構成することであるが、誤差の推定が多段階で必要であり、簡単な計算では求めにくい。そこで本論文では標本誤差を評価するためにモンテカルロ・シミュレーション

を適用する。もちろん現実の議席予測では非標本誤差も含まれている。本論文で示す議席の予測区間は標本誤差だけを見積もった結果であるが、現実の議席予測にとっては「そこまでは達成可能な到達すべき水準」という目標の設定値としての役割を与えるものだと考えられる。またモンテカルロ・シミュレーションによって構成した予測区間と、実際のマスコミ各社が公表した予測区間との差分が明らかになるので非標本誤差の大きさの程度に関する知見も得られる。

## 2. 選挙予測の性質

選挙予測のためのデータ作成手段である標本調査は、形式的には世論調査と同じである。しかし選挙予測調査に特有の相違点もいくつかあるので、本節で再確認する。なお、マスコミの選挙予測には二種類ある。投票日の約一週間前に有権者に対して実施する標本調査による予測と、投票当日に投票者に対して実施する標本調査による予測である。前者は主に新聞社が実施・報道するもので世論調査と同じ形式である。後者は主にテレビ局が実施・報道するもので、「出口調査」と呼ばれている。本論文が扱う選挙予測は前者である。

選挙予測調査に特有の性質の第一は、母数が明らかになることである。一般的な世論調査では母数は未知のままである。たとえば世論調査にとって最も重要な測定項目である内閣支持に関して、我々は標本比率を得ることができるが、母集団の真値（母数）は未知である。質問文を変えると「内閣支持」の標本比率も変化するという意識測定方法上の問題もあり、このことを林(1993)は「調査は実態ではない」と表現した。世論調査の標本観測値は、母数が未知のまままだという意味において結果が「正しい」のか「正しくない」のかを問うことができない。しかし選挙予測調査は違う。投票行動の結果（母数）は開票後に明らかになる。本研究でモンテカルロ・シミュレーションを適用できるのはこの性質のためである。

第二の性質は、標本調査のための母集団を固定できないことである。世論調査の母集団は「ある時点の有権者」と定義して有権者名簿を抽出枠に使用できる。しかし選挙予測調査の目標母集団は「一週間後の選挙の投票者」である。それは「有権者」の部分集合ではあるが、特定できない。母集団のこの不定性は投票率の低下に応じて深刻な問題になる。最近三回の総選挙（第40回～第42回）投票率の平均値は63%である。目標母集団（投票者）の抽出枠からの標本抽出が原理的に不可能である以上、投票日前に標本抽出をする選挙予測調査では、目標母集団を部分集合として含む「有権者名簿」を抽出母集団とするしかない。母集団と標本の関係は、世論調査とも出口調査とも異なる。選挙予測調査の精度が出口調査より悪い原因の一つはここにある。出口調査は「投票者」という目標母集団に対する標本調査なのである。

第三の性質は、標本誤差と非標本誤差を分離できることである。母数が明らかになるという第一の性質のために、母数と標本観測値の差（全誤差）が算出できる。全誤差は標本誤差と非標本誤差の和であると定義すれば、非標本誤差が標本誤差より大きい場合は、全誤差と標本誤差の残差を非標本誤差の大きさであると解釈することができる。本研究では、標本誤差をシミュレーションで示す。現実の選挙予測は非標本誤差も含みながら過去の新聞紙面に示されている。

以上の性質は、統計的推測の適用だけでは選挙予測が困難であることを示唆している。伝統的なネイマン・ピアソン流の推測統計学の理論は、標本からの推定値に関して頻度論的な信頼区間を構成するが、母数は定数で動かないという前提がある。しかし選挙予測の母集団は投票日まで不定であり、その意味で動的である。また統計理論が想定する標本は、母集団から回収率100%で無作為抽出された大きさ  $n$  の標本に関する議論をしており、非回収・無回答を考慮しない。しかし選挙予測は「投票態度未定者」という無回答も含めて予測しなければならない。統計学は標本誤差に関して理論を完成しているが、非標本誤差について一般的な方法を提供していない。しかし選挙予測の課題は非標本誤差に関する具体的な分析方法である。

### 3. データ

衆参両院の国政選挙のうち衆議院議員選挙（総選挙）の投票結果データを分析する。対象とする総選挙は第

二次世界大戦後の現憲法下で実施された第23回（1947年）総選挙以降のうち、保守合同による自由民主党（自民党）結成と社会党左右合同による日本社会党（社会党）結成に象徴される「55年体制」が確立されたあと実施された、第28回（1958年）総選挙から、第42回（2000年）総選挙までの15回分とする。

第40回（1993年）までは中選挙区単記投票制だが、第41回（1996年）と第42回（2000年）は小選挙区比例代表並立制に、選挙制度が変更された。そこで第41回と第42回に関しては小選挙区（300議席）のみを比較対象として扱い、比例代表に関しては分析対象から除外した。

分析データとしては主に水崎節文による総選挙データベース「JED-M」を利用した。ここには選挙区別に全候補者の政党、得票数などが収録されている。

### 4. シミュレーションの方法

第  $g$  回総選挙の選挙区  $h$  における、 $k$  人の各候補者の得票率（母比率）を  $P_j$  と書くと、

$$\sum_{j=1}^k P_j = 1 \quad (1)$$

である。候補者を一定の順番（たとえば得票数順）に並べた時の第  $j$  候補の累積得票率を  $T_j$  とする。

$$T_j = P_j + T_{j-1} \quad (2)$$

ただし、 $T_0=0$  である。

母集団（第  $g$  回総選挙の選挙区  $h$ ）から標本サイズ  $n$  の単純無作為抽出をシミュレーションするために、乗算合同法で擬似一様乱数列を  $n$  個発生させる。すなわち、漸化式

$$X_n = aX_{n-1} \pmod{M} \quad (3)$$

を用いて非負整数列  $\langle X_n \rangle$  を生成し、一様分布  $U[0,1)$  に従う擬似乱数列  $x_n$  を得るために、

$$x_n = \frac{X_n}{M} \quad (4)$$

と変換する。

$i$  人目の模擬投票者に与えた乱数  $x_i$  に対して、候補者を  $T_1$  から順番に比較し、判定式

$$x_i < T_j \quad (5)$$

が真となった候補者  $j$  が得票したとする。



$n=400$  では、第一党が 12~17 議席、第二党が 8~16 となり、標本サイズが倍になることで、第一党の予測区間は 2~3 議席程度狭くなることが分かる。

標本誤差だけしか考慮しなかったにもかかわらず、予想に反して、95%信頼区間に選挙結果（母数）を含まない総選挙が 15 回のうち 3 回あった。選挙結果が予測区間に含まれなかったこの現象を「予測の失敗」と呼ぶことにする。予測の失敗は、佐藤内閣の二度目の総選挙（第 32 回；1969 年）、大平内閣の二度目の総選挙（第 36 回；1980 年）、中曽根内閣の二度目の総選挙（第 38 回；1986 年）の三回である。表 1 では区間の上下限に下線を引いて「予測の失敗」を示したが、標本サイズが 200 でも 400 でも同じ三回の総選挙で予測が失敗している。第 32 回（1969 年）では、 $n=200$  で第二党の予測にも失敗している。

失敗した予測は三回とも第一党の獲得議席を過小評価し、95%信頼区間の上限が結果に届かなかった形である。つまり自民が予想以上に勝った。擬似乱数のシード値をいろいろ変更して再度シミュレーションしてみたが、結果は同じであった。試みに標本サイズを  $n=1000$  と  $n=2000$  に増やしてシミュレーションしたところ  $n=1000$  ですべての総選挙が 95%信頼区間に選挙結果が含まれた。

## 6. 戦後のマスコミ選挙予測の結果

それでは、マスコミ（新聞社）による現実の選挙予測結果はどうであったのか。表 2 に朝日新聞社、毎日新聞社、読売新聞社の三社の紙面から、議席予測の結果を抜き出してまとめた。

### 【このあたりに表 2】

第 28 回（1958 年）総選挙は、「55 年体制」が確立して以後はじめての総選挙であると同時に、マスコミ（朝日新聞社）の選挙予測が初めて全選挙区対象の標本調査を実施する体制を整えた時でもあった。また予測手法に関しても、朝日新聞社は第 28 回（1958 年）総選挙から、林・高倉(1964)で示された統計的データ解析（主に数量化 1 類と 2 類）を適用して議席予測を始めたのであった。毎日新聞社と読売新聞社が同様の「科学的方法」で選挙予測報道を始めたのは、第 30 回（1963 年）総選挙からである。

朝日新聞社では政党別獲得議席の予測区間を発表し

た 11 回のうち、第一党の議席数に関して 6 回の予測失敗があった。毎日新聞社では 13 回のうち 5 回、読売新聞社では 6 回のうち 3 回が失敗であった。各社とも半分程度は失敗していることになる。ただしこれは予測区間を発表した分についてのみである。朝日新聞社では、第 35 回（1979 年）の予測の失敗から 3 回連続で失敗をしたあと、第 38 回（1986 年）から第 41 回（1996 年）まで 4 回の総選挙において予測区間を発表せず、当選確率の水準別人数をグラフで示す報道形式を採用した。この 4 回の総選挙では区間予測が成功している可能性がある。読売新聞社もグラフ表現による報道が多く、第 40 回（1993 年）以降はグラフ表現を継続している。毎日新聞社だけが一貫して第 30 回（1963 年）以来、区間予測を公表している。

## 7. 考察と展望

マスコミによる現実の予測と、シミュレーションによる予測を比較すると、当然ではあるが現実の予測成績の方が悪い。予測区間に関して、最近時点の第 42 回総選挙（2000 年）で比較してみると、シミュレーションでは、レンジが 14 議席（ $n=400$ ）であるのに対して、朝日新聞社が 23 議席、毎日新聞社が 24 議席と 10 議席近く拡大している。しかも両社とも予測に失敗している。自民の議席を過大に予測したのであるが、シミュレーション結果を見ても、選挙結果の議席数は分布の中央付近ではなく下限に近いところにあるので、標本誤差だけから予測しても難しい選挙であったといえる。

シミュレーションにおいてさえ予測に失敗した三回の総選挙は、現実の予測でも予測区間を発表した全社が失敗している。内容も自民の勝利を過小評価するという同形の失敗となっている。なぜ三回の総選挙は、標本誤差しか考慮していないシミュレーションにおいてさえ失敗するのか、という問題は本研究の過程で新しく生じた課題である。

第 32 回（1969 年）の自民の得票率を、同じ佐藤内閣の前の総選挙（第 31 回；1967 年）と比較すると、自民党の得票率はほとんど同じである。むしろ 48.80% から 47.63% にやや低下したにもかかわらず議席を増やして圧勝したのである。この二つの総選挙が、他の選挙と比較してどのような構造的相違を内包しているのかを定式化するのは今後の課題としたい。試みに選挙間の接戦度を指標化して比較してみたが注目すべき

特徴差を発見できなかった。また、接戦が多ければ当落の判別力に相違があるだろうとの仮説から、佐藤、大平、中曽根の三内閣が実施した二回ずつの総選挙で、得票率と定数を説明変数とし、当落を目的変数とする二群の線形判別分析を適用してみたが、判別率に大きな違いはない。表3は三内閣のうちから佐藤内閣の二つの総選挙に対する交差妥当化判別分析の結果を示したものである。第31回と第32回の正判別率はそれぞれ、88%と89%で大差ない。

表3 第31, 32回総選挙の交差妥当化判別率

31回	当選	落選		32回	当選	落選	
当選	474	12	486	当選	473	13	486
落選	97	334	431	落選	92	367	459
	571	346	917		565	380	945
正判別率=88%				正判別率=89%			

## 8. おわりに

マスコミの選挙予測は、第一段階として、各候補者の調査支持率から推定得票率（点推定値）を求め、第二段階として各候補者の当選確率を考慮しながら政党別獲得議席数の予測区間を構成するという手順が一般的である。第一段階は非標本誤差の分析であり、第二段階は標本誤差の分析である。そして第一段階の方が困難であり、重要であり、各社のノウハウが蓄積される分野である。最初の得票率の推定（点推定）に構造的な偏りがあると、後に続く区間予測が失敗するからである。

本論文では、第一段階には言及せずに、第二段階の標本誤差についてのみ検討したが、モンテカルロ・シミュレーションで示されたように、予測に失敗する選挙が存在した。それが、どのような支持分布の構造を持つかは明らかではないが、標本誤差だけであっても、予測しやすい選挙と予測困難な選挙が存在するという知見を得た。

現憲法化で実施された戦後の総選挙のうち、最初の第23回（1947年）から第27回（1955年）までの5回については分析しなかった。REEDによるデータベース「衆議院議員総選挙全候補者別得票結果 1947年～1993年」があるのだが、所属政党や得票数に間違いが多い。戦後初期5回の総選挙については筆者が修正したが、最終的に本研究でREEDの選挙データを分

析することは断念した。

## 参考文献

- [1] 水崎節文、『総選挙データ・ベースの開発とその利用』、エル・デー・ビー。
- [2] 伏見正則、『乱数』、東京大学出版会、1989。
- [3] 津田孝夫、『モンテカルロ法とシミュレーション 電子計算機の確率論的応用』、培風館、1995。
- [4] 小西貞則、「ブートストラップ法による推定量の誤差評価」、『パソコンによるデータ解析』、朝倉書店、1988。
- [5] 林知己夫・高倉節子、「予測に関する実証的研究 選挙予測の方法論」、『統計数理研究所彙報』、12,1、1964。
- [6] 林知己夫、『数量化の方法』、東洋経済新報社、1974。
- [7] 林知己夫、「世論調査の発展と現状」、『世論調査の現状と課題』、至誠堂、1977。
- [8] 林知己夫、『行動計量学序説』、朝倉書店、1993。
- [9] 鈴木督久、「世論調査：日経・朝日・読売の内閣支持率比較」、『日経リサーチ・レポート』、1999-I、1999。
- [10] 鈴木督久、「内閣支持率とその表現」、『日経リサーチ・レポート』、2000-III、2000。
- [11] 鈴木督久、「選挙予測の宿命」、『日経リサーチ・レポート』、2001-III、2001。
- [12] 鈴木督久、「第19回参院選・追跡調査報告」、『日経リサーチ・レポート』、2001-IV、2001。

表1 1958年以降の総選挙における議席獲得シミュレーション

総選挙 回 年	内閣	定数	獲得議席数		獲得議席数の95%区間予測			
			第一党	第二党	n=200		n=400	
					第一党	第二党	第一党	第二党
28 1958	岸	467	287	166	282 (18) 300	152 (18) 170	283 (16) 299	155 (8) 170
29 1960	池田	467	296	145	287 (17) 304	132 (15) 147	290 (15) 305	134 (14) 148
30 1963	池田	467	283	144	271 (18) 289	134 (17) 151	274 (16) 290	135 (15) 150
31 1967	佐藤	486	277	140	260 (18) 278	135 (17) 152	263 (16) 279	136 (16) 152
32 1969	佐藤	486	288	90	<u>264</u> (18) <u>282</u>	<u>91</u> (19) <u>110</u>	<u>271</u> (15) <u>286</u>	89 (16) 105
33 1972	田中	491	271	118	254 (18) 272	110 (15) 125	258 (16) 274	111 (15) 126
34 1976	三木	511	249	123	237 (18) 255	110 (16) 126	240 (16) 256	111 (15) 126
35 1979	大平	511	248	107	240 (19) 259	99 (16) 115	241 (17) 258	100 (15) 115
36 1980	大平	511	284	107	<u>266</u> (15) <u>281</u>	100 (15) 115	<u>271</u> (12) <u>283</u>	101 (14) 115
37 1983	中曽根	511	250	112	246 (17) 263	103 (14) 117	245 (17) 262	104 (13) 117
38 1986	中曽根	512	300	85	<u>280</u> (15) <u>295</u>	83 (15) 98	<u>285</u> (13) <u>298</u>	82 (14) 96
39 1990	海部	512	275	136	262 (16) 278	127 (10) 137	265 (15) 280	129 (9) 138
40 1993	宮沢	511	223	70	208 (17) 225	66 (17) 83	212 (15) 227	65 (14) 79
41 1996	橋本	300	169	96	154 (19) 173	88 (18) 106	157 (17) 174	89 (16) 105
42 2000	森	300	177	80	174 (18) 192	67 (16) 83	175 (15) 190	69 (14) 83

注1:1996年以降は小選挙区比例代表並立制. 議席数は小選挙区のみを対象とした

注2:政党別の議席数は選挙時点の所属で分類してある

注3:第一党は自民,第二党は社会.ただし1996年,2000年の第二党は新進,民主

注4:推定区間の中間にある括弧内の数字は予測区間のレンジ

注5:下線を引いた予測区間は選挙結果を含まなかったことを示す

表2 投票日の約1週間前の標本調査による予測議席数の推定区間

総選挙 回 年	朝日新聞社		毎日新聞社		読売新聞社	
	第一党	第二党	第一党	第二党	第一党	第二党
28 1958	<u>280</u> (6) <u>286</u>	<u>172</u> (6) <u>178</u>	取材調査		政党支持	
29 1960	285 (20) 305	<u>128</u> (4) <u>142</u>	取材調査		取材調査	
30 1963	280 (18) 298	<u>146</u> (14) <u>160</u>	283 (15) 298	<u>147</u> (13) <u>160</u>	<u>287</u> (10) <u>297</u>	<u>149</u> (6) <u>155</u>
31 1967	263 (16) 279	133 (16) 149	268 (10) 278	<u>149</u> (8) <u>157</u>	267 (10) 277	<u>147</u> (6) <u>153</u>
32 1969	<u>267</u> (16) <u>283</u>	<u>110</u> (16) <u>126</u>	<u>272</u> (7) <u>279</u>	<u>118</u> (10) <u>128</u>	取材調査	取材調査
33 1972	269 (20) 289	101 (18) 119	271 (18) 289	112 (13) 125	取材調査	取材調査
34 1976	246 (20) 266	106 (18) 124	244 (20) 264	122 (16) 138	247 (16) 263	<u>128</u> (15) <u>143</u>
35 1979	<u>260</u> (20) <u>280</u>	93 (18) 111	<u>262</u> (12) <u>274</u>	102 (20) 122	<u>266</u> (13) <u>279</u>	<u>112</u> (13) <u>125</u>
36 1980	<u>263</u> (18) <u>281</u>	96 (18) 114	<u>256</u> (19) <u>275</u>	95 (24) 119	G	G
37 1983	<u>267</u> (22) <u>289</u>	<u>88</u> (18) <u>106</u>	250 (30) 280	101 (19) 120	<u>256</u> (26) <u>282</u>	95 (28) 123
38 1986	G	G	<u>265</u> (28) <u>293</u>	85 (27) 112	G	G
39 1990	G	G	254 (24) 278	133 (9) 142	245 (33) 278	133 (10) 143
40 1993	G	G	211 (20) 231	59 (19) 78	G	G
41 1996	G	G	166 (23) 189	<u>69</u> (23) <u>92</u>	G	G
42 2000	<u>182</u> (20) <u>202</u>	<u>53</u> (18) <u>71</u>	<u>188</u> (23) <u>211</u>	<u>48</u> (24) <u>72</u>	G	G

注1:「取材調査」は標本調査をせず記者の取材による予測報道。「政党支持」は議席予測せず支持率のみ報道

注2:「G」は、議席の予測区間を公表せず、当選確率の水準別の人数をグラフで示した報道

注3:推定区間の中間にある括弧内の数字は予測区間のレンジ

注4:下線は選挙結果を区間に含まないことを示す。選挙年の下線はシミュレーションでも予測に失敗していることを示す

注5:第一党は自民,第二党は社会。ただし1996年,2000年の第二党は新進,民主