

シミュレーション人口学：入門以前

中澤 港(東京大学大学院医学系研究科人類生態学教室)

人口学でシミュレーションといえば、ふつう、人口モデルの数値計算による予測あるいは過去の推計をさす。これらは、メカニズムよりも数値間に現れる関係式に基づいてアприオリに立てたものが多く、観測不可能な変数も多い。この推計においてもっとも重視されるアウトプットは、幅はついているにせよ「推計値」である。この種のシミュレーションを Dyke (1981)は、「marginal use (訳してみれば『枠外の使用法』とでもなるう)」と呼んで批判している。モデルが正しいかどうかを評価する手段がないのに推計値に意味をもたせるのは危険だというのである。では、人口学におけるシミュレーションの正当な用法とは何になるだろうか。同じく Dyke によれば、次の2つに集約される。

一つは、シミュレーションによるデータの生成である。本質的に観測不可能な事象でも、シミュレーションによってデータを生成すれば、完全なデータセットを得ることが可能である。例をあげよう。無文字社会における過去の人口復元のために使える手段は、ふつう、聞き取りによる家系図の再構成である。その上で復元された人口がどのように変化してきたかを解析する。しかし、データとして信頼できる聞き取りが可能なのはたかだか 50~100 年程度であり、対象者がどんなにすばらしい伝承能力をもつ人々で、かつ研究者がどんなにすばらしい聞き取り能力をもつとしても、200 年以上にわたる個人レベルの詳細な人口動態を完全に復元することは不可能である。したがって、信頼性のないデータが解析モデルにどういったバイアスを与えるかを評価する必要がでてくる。こういう場合にシミュレーションを使えば、出生、死亡などについていろいろなレベルを設定した上で、完全な人口動態データを短時間のうちに得ることができる。この完全なデータに対して解析モデルを適用して得られる結果と、何らかの脱落を仮定して得られるデータに対して同じモデルを適用して得られる結果がどの程度一致するかによって、聞き取りの不完全さに対する解析モデルの頑健性を評価できるわけである。さらに、乱数の初期値だけを変えて何度もシミュレーションを実行すれば、与えた初期条件の下でのデータの分布を得ることができる。これは一般にコンピュータ集約型統計学と呼ばれる方法で、解析的に求めにくい分布関数を、シミュレーションならば容易に求めることができる。

もう一つは、単純な解析的モデルが存在しないような、複雑な現象についてシミュレーションを用いることである。たとえば、人口再生産過程について、出生、死亡、結婚を分けてモデルを構築することは可能だが、現実¹に即した結婚過程を解析的に表現することはまず不可能である。なるほど、稲葉(1997)が指摘するとおり、Fredrickson (1971)以来、男女のペア形成の分布を考慮した非線形微分方程式モデルはいくつか提案されているが、いくつかの仮定のもとで指数関数的成長解の存在が示せたなどの成果はあるものの、推定の誤差も大きく、かつそれと現実の婚姻規制との関連が「年齢間競合の公理」など基本的な点をのぞいては明示されていない。たとえば、交叉イトコ婚を数学的に表現するには、まず交叉イトコという親族関係の変数を作り、年齢構造とは別にモデルに組み込んで、親族関係ごとに関数を定義しなければならないわけで、この手口ではきわめて難しい。その点、個人レベルのシミュレーションならばルールベースという形で結婚を表現することは容易である。そのルールベースが正しいという保証がないのが本質的な弱点であるが、それ以外の方法ではまったく解が得られないような場合なら、シミュレーションの利用は大いに価値がある。個人レベルの単純なルールを決めると全体の動きがうまく表現できるという意味では、生態学の Individual-based model におけるシミュレーションの利用法(DeAngelis and Gross, 1992)と同等である。MacCluer らのグループによって 1970 年前後からおこなわれた一連の研究は、まさにこの意味でシミュレーションを利用していたが(MacCluer *et al.*, 1971; MacCluer, 1980) その後はあまり進展しなくなった。近年では、杉藤(1991)によるヨロongoの複雑な婚姻規制のシミュレーションが見事な例であるが、あまり報告が多いとはいえない。

「シミュレーション人口学」は、上記2つの目的でおこなうシミュレーションを含む人口分析をさす、私の造語である。ペア形成を含めて人口再生産のシミュレーションを個人レベルでおこなうことは、2つの目的のどちらにも利用でき、シミュレーション人口学の基本といってよい。そこで本稿では、個人レベルの人口再生産シミュレーションについて説明することにする。

個人レベルの人口再生産シミュレーションの枠組み

個人レベルで人口再生産を表現するには、おおまかにいえば、(1)「個人」を定義する、(2)死亡の発生確率を決める(年齢の関数とすることが多い)、(3)ペア形成のル

ールを規定する, (4)ペアの出生力を決める, の4段階が必要である。しかる後に, それらを組み合わせることによって, 集団全体の人口再生産がおこなわれる。ただし, Goldman らの一連の研究が既婚未婚が死亡に影響を与えることを示唆しているので (たとえば Hu and Goldman., 1990), 死亡を個人レベルのイベントとするこの枠組み自体も検討の余地を残している。

実装としてはコンピュータ上のプログラムとすることが一般的であり, 確率事象の発生は擬似乱数を使って表現される。プログラム言語としては, かつてはライブラリが充実していることから FORTRAN が使われることが多かったが, 現在では共通のライブラリがフリーで使えるので, 自分の使いやすい処理系を選べばよい。ただし, 個人を定義する必要上, C++や Java のクラス, あるいは C の構造体や Pascal のレコード型のような変数の型を宣言できる言語が適していると思われる。シミュレーション専用の言語を使うこともできるが, パーソナルコンピュータでもワークステーションでもメインフレームでもスーパーコンピュータでも使えるという点から考えると C や Fortran の方が便利であり, 筆者は C を使っている。なお, どの段階でも単位時間を決めることは重要である。世代を単位時間とするのか, 年か月か週か日かというのは, 大きな違いである。

かつては, ペア形成を考えない, 単性のシミュレーションもよくおこなわれていた。なぜなら, 人口再生産はほぼ女性のみによって決まるとする考え方が一般的だった (Wood, 1994)からである。しかし近年では, 男性の寄与も小さくないことが明らかになっており, 単性のモデルでは人口再生産をあらわすのに不十分だとわかってきた。最近マスコミを騒がせている, 「環境ホルモン」による精子数減少や精巣ガンの増加と少子化の関連の指摘 (Skakkebaek et al., 1998) ばかりでなく, 育児に対する男性の意識や行動の違いがカップルの出生力と関係している (Gohel et al., 1997) といった社会経済的側面についての指摘もある。それゆえ, ペア形成を考えることには大きな意味がある。

「個人」の定義

「個人」を定義するとは, ヒトの属性のうち, どの部分をピックアップするかということである。最低限必要なのは, 個体識別子, 親子関係, 年齢 (月齢, 日齢), 配偶関係であるが, 居住地や疾病抵抗性遺伝子といった属性を扱うこともできる。どの属性を要素

に含めるかは、死亡やペア形成のルールベースを決める際の必要性から決定する。シミュレーションの実装を C 言語でおこなう場合、個人は構造体として定義するのが自然である。たとえば、下記のようなコーディングができる(なお、C 言語そのものについては本稿では説明するスペースがないので、成書(たとえばカーニハン&リッチー, 1989)を参照されたい)。

```
typedef struct individual
{
    struct individual *next;          /* the pointer to next person */
    struct individual *spouse[MAXNUMSPOUSE]; /* the pointer to the
spouse */
    unsigned long idnumber; /* idnumber */
    unsigned long pspouseid; /* primary spouse's idnumber */
    unsigned int generation; /* generation */
    unsigned int numchild; /* number of children */
    unsigned int clan; /* clan */
    unsigned int numspouse; /* number of spouses */
    unsigned int gender; /* male or female */
    unsigned int age; /* age */
    unsigned int region; /* region, risk of mosquito bites */
    unsigned int allele[2]; /* now population */
    unsigned int marriage; /* marriage status */
    unsigned int infect; /* infected by malaria or not */
} idv;
```

この構造体 `idv` へのポインタの連鎖として集団全体を定義することができ、さらに全成員の挙動を各単位時間において決定し、属性値を変えることで、シミュレーションが進行する。上の例では属性値を数値のまま扱っているが、列挙型を使って見やすくし、ビットフィールドを使うなどしてメモリ使用量を減らす工夫をする方が良い。

死亡確率の決定

次は、全成員の挙動を各単位時間において決定する。生存時間解析の加速モデルでは死亡ハザードや生存関数に単純な分布を仮定し、生存時間データを使って最尤法でパラメータを求めるが、逆にある分布に対してパラメータを与えれば生存時間分布が計算できるので、それにしただがって各個人の死亡が起こるようにシミュレーション条件を設定することができる。ハザード解析では、指数分布あるいはワイブル分布が用いられることが多いが、シミュレーションでは分布を表の形で与えてもよいので、「モデル生命表」(たとえば、Coale and Demeny, 1983)もよく用いられる。

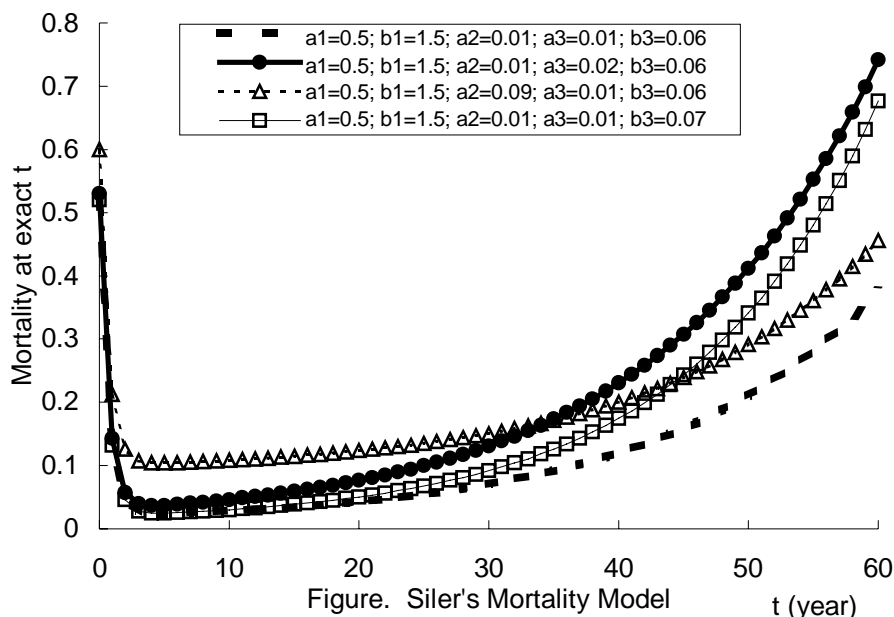
いくつか「意味のある」パラメータを与えて年齢別死亡確率(コホートで見れば生存時間分布)を決めるためには、ゴンペルツ曲線など古くからいろいろ提案されてきた死亡曲線の考え方が有効である。なかでも、死亡のメカニズムを十分に考慮したモデル

は、Siler の 5 パラメータモデル以降とってよい。Siler モデルは、乳幼児死亡が高いことと、成人してからは年齢とともに死亡リスクが上がってゆくことを考慮したものである (Mode and Salsberg, 1993)。年齢別 (正確な時刻 t における) 死亡ハザード h_t が次の式で表される。

$$h_t = a_1 e^{-b_1 t} + a_2 + a_3 e^{b_3 t}$$

右辺第一項は、幼児期の間、年齢とともに減少する死亡率を示している。この減少パターンは、出生の瞬間に未成熟であることからくる死亡の大きさを表すパラメータ a_1 と、小児期の減少のしかたを示すパラメータ b_1 によって決まる。第二項は、年齢と関係なくランダムに起こる死亡をあらわす Makeham の定数 a_2 である。第三項は、加齢に伴う死亡の増加を示すゴンペルツ成分である。 a_3 は出生の瞬間に決まるこの成分からくる死亡の初期値の大きさを示し、 b_3 は加齢にともなう死亡の増加率の大きさを示す。ゴンペルツ成分は、死亡というものを、生命体を構成する複数の部品が順次壊れていき、ある程度故障がたまったところで起こると考えれば納得できる。Siler モデルではいくつかの国レベルの生命表に対してあてはめを行って得られたパラメータも発表されているので、シミュレーションのベースラインを決めるのが容易である。

Siler モデルから得られる死亡曲線の例を図示した。 t に対して乗算になるため、ゴ



ンペルツ成分の増加パラメータ b_3 のわずかな変化にも死亡曲線が敏感に変化している。Siler モデル以降、パラメータを増やして実人口の死亡パターンへの適合度を上げるための努力はいろいろされてきており、10パラメータの Mode-Jacobson モデルなどかなりあてはまりはよいが、反面、複雑化しすぎているために個人レベルのシミュレーションを行うには向いていない。

別のアプローチとして、老年期における生活習慣病などの死亡の増加を、加齢自体が「相互に依存し合った故障の滝のようなもの」によって引き起こされると考えて説明する、「雪崩モデル」も提案されている(Gavrilov and Gavrilova, 1991)。この考え方は、本質的に個人ベースのシミュレーションに馴染むものであり、年齢別死亡ハザードという形でなく、より自然に死亡をモデル内にとりこむことができる。 λ_0 をランダムに蓄積する加齢効果のパラメータとし(つまり年齢が増えるとともに死亡ハザードが λ_0 だけ増える)、 λ が一つの故障から「滝」が誘導される確率とする。さらに、故障の「滝」によって起こる死亡のハザードが故障数に比例する(つまり、 n 個の故障をもつ個人についての死亡のハザードが $n\mu$ である)とするわけである。シミュレーションにはこのままの形で導入できるが、この場合の死亡ハザードの式を考えてみよう。 $t = 0$ での故障数が 0 であり、故障数に上限がないとすると、死亡の総ハザードは、

$$h(t) = \frac{\mu\lambda_0\{1 - \exp[-(\lambda + \mu)t]\}}{\{\mu + \lambda \exp[-(\lambda + \mu)t]\}}$$

となる。ここで $\lambda \gg \mu$ であれば、この式は Siler モデルの第2項と第3項を合わせたものに縮退するので、乳幼児死亡を別に考えさえすれば、このモデルで得られる年齢別死亡ハザードがもっともらしい形になることは保証される(注: Wood *et al.*, 1992 にはこう書かれているが、筆者は原著を読んでいないので式変形の確認はしていない)。

一方、マクロな人口モデルでは2つという少ないパラメータで any population の年齢別死亡確率を表せる Brass (1975)のロジットシステムもよく使われており、これも個人レベルの死亡スケジュールとして利用できる。簡単に言えば、基準となる年齢別死亡率を決め、2つのパラメータ α と β でその形を変える方法であり、計算が簡便なこととパラメータの意味(実際の集団にあてはめたときの値を含む)がよくわかっているのが利点である。 α によって全体の死亡水準、 β によって乳幼児期と成人してからの死亡水準の比を表現することができる。しかし、大規模な人口の生命表との適合度で検証した

結果によれば、Siler モデルや Mode-Jacobson モデルに比べると、かなりあてはまりは悪いといわざるをえない。人口研究に使われた例はまだ少ないが、今後は雪崩モデルのようなアプローチが有望と思われる。

ペア形成のルール

解析的なモデルではペア形成は2変量間の結婚関数として定義する必要があるが、シミュレーションではそういう形になりにくい複雑な関係も、条件式のかたまりとして設定できる。たとえば、

```
if(((sspouse->age)<=spouseage+SPOUSEAGEDIFHMAX) &&
((sspouse->age)>=spouseage-SPOUSEAGEDIFLMAX) &&
(sspouse->gender==MALE) &&
(sspouse->numspouse<nmaxnumspouse) &&
(qs->region==sspouse->region) &&
(qs->clan!=sspouse->clan)) {
    list->next=lastlist;
    lastlist=list;
    list->pointer=sspouse;
}
```

というように、年齢差何歳未満で、異なるクランに属していて、配偶者が何人以内といった条件を満たす配偶者を探して配偶者候補リストに入れるというアルゴリズムは、わずか数行で記述できる。

一方、解析的なモデルの結婚関数を条件式のかたまりにすることもペア形成は実現される。結婚関数は、ふつう、二変量の確率密度関数である。たとえば、Farlie-Morgenstern 関数(Mode and Salsberg, 1993)の例がある。このモデルでは、まず対数正規分布を使った「結婚市場への加入」を考え、結婚市場の男女のカップリングが Farlie-Morgenstern 関数にしたがう、という2段階のモデル化を行っている。これを条件式に書き換えるのは難しくないが、パラメータの意味づけが難しい。

ペアの出生力の決定

次は、ペア単位で出生力を決める。出生力を規定する要因の分解としては、Bongaarts (1978)の近成要因モデルが有名である。これは、出生力に直接影響する要因を、曝露要因(既婚者割合)、意図的な有配偶出生力制限要因(避妊及び人工妊娠中絶)、及び有配偶自然出生力要因(授乳性無月経、性交頻度、不妊、子宮内自然死亡、妊娠可能期間)として、間接的に影響するその他の要因とは区別して考える

ものである。有配偶出生力については、Coale and Trussell (1974)のモデル出生スケジュールが有名である。しかし、出生力を年齢別既婚者割合と年齢別有配偶出生力の積の和とする考え方の問題点は少なくともふたつある。ひとつは、「年齢別」出生力という時間軸の切り方である。年齢だけでなく結婚後の年数も明らかに出生力には影響するし、出生コホートによる違いもありうる(たとえば国立社会保障・人口問題研究所, 1998)。もうひとつの問題点は、女性についてしか考えられていないことである。かつて男性は一生精子供給能力をもつとされてきたが、男性側の出生力もライフステージによって変わることはいまや常識である。

したがって「ペアの出生力」というものを考えねばならないが、この問題についての研究は皆無に近い。これまで行われたシミュレーションではそれを理由にして女性の年齢だけを出生力を規定するパラメータとしているものが多い(たとえば, Mode and Salsberg, 1993)。一つの案としては、先にあげた Bongaarts の近成要因モデルに男性側の要因も追加することも考えられるが、これをきちんと定式化するのはきわめて難しく、今後の課題である。

乱数の生成

データラメに選び出した数を乱数といい、その列びを乱数列というわけだが、そもそも「乱数」とは何かと改めて考えてみると、数学的に厳密に定義することはきわめて難しい。Knuth (1981)にも、「この話題について思考をめぐらす人々は、ほとんど必然的に“乱数”とは何かという哲学的な議論にのめり込んでしまう」と書かれている。その上で、Knuth は、次のように論を進めている。「まず、1個の乱数などというものは存在しない。たとえば、2は乱数であろうか。そうではなくて、われわれは、ある特定の“分布”をもつ“独立な乱数の系列”を扱う。その大体の意味は、各数が単に偶然によって選ばれており、系列中の他の数とは何の関係もなく、各数が任意の区間に落ちる確率が定まっている、ということである。」さらに、Lehmer の定義「乱数列とは曖昧な概念であって、説明を聞かないものにとっては各項を予測することが困難であり、各数字は統計学者になじみの検定のうち、利用目的に応じて選んだいくつかのものに合格しているような数列である」と Franklin の定義「数列(1)が、一様分布にしたがう確率変数の独立な標本の無限系列がもっているすべての特徴をもっているならば、でたらめであるという」を引

用して, Franklin の定義を厳密に適用すると乱数列というものは存在しなくなるので, Lehmer の定義を精密化する, といって, 次の2つの定義を与えている。

定義 R5: b 進数列より計算可能な部分列規則により定まる部分数列がすべて1次均等分布であるならば, この b 進数列を“乱数列”と定義する。 $[0,1)$ 数列 $\langle U_n \rangle$ から作られる b 進数列 $\langle \lfloor bU_n \rfloor \rangle$ がすべての整数 $b \geq 2$ に対して“乱数列”であるならば, $\langle U_n \rangle$ を“乱数列”と定義する。

定義 R6: b 進数列 $\langle X_n \rangle$ にたいして, 相異なる非負整数 s_n の無限数列を, n と $X_{s_0}, \dots, X_{s_{n-1}}$ の関数として定める任意の実効的算法を適用したときに, 得られる部分数列 $\langle X_{s_n} \rangle$ が定義 R5 の意味で“乱数列”であるときに, $\langle X_n \rangle$ を乱数列と定義する。 $[0,1)$ 数列 $\langle U_n \rangle$ から作られる b 進数列 $\langle \lfloor bU_n \rfloor \rangle$ がすべての整数 $b \geq 2$ に対して, 上の意味で“乱数列”ならば, $\langle U_n \rangle$ を乱数列と定義する。

明らかに, この定義によれば有限乱数列というものは存在しないので, Knuth は「有限乱数列」の定義も与えているが, それは算法の集合依存であり, 実用的でない。また, Wald 数列のようなものも計算時間がかかりすぎるので実用的でない。したがって, コンピュータでは, 「乱数列らしく見える」という意味での擬似乱数列しか生成できないのが現状である。

しかし, シミュレーションで用いるには, 次の4つの性質を満たしていれば, 擬似乱数列で十分である。(1)連続する値の間に線形の関係がない, (2)値の出現する期待値がとりうる値全部について等しい, (3)短い周期の繰り返しにならない, (4)算法で生成される。つまり, 「ある確率で事象が起こる」ことをあらわせばよい。確率0.3で事象Aが起こるなら, 100000回シミュレーションをしたときに 30000 ± 1500 回くらいAが起こってくれて, 起こるのと起こらないのが適度に散らばっていればよいのである(ただし, 実験と同じで, 条件を同じにした場合に結果が再現できなくてはいけないので, 実行時刻を変換するとか, 計算時間を使うとかいうのは駄目である。あくまで, 算法でなければならない)。

この意味でよく用いられる擬似乱数列生成アルゴリズムは Lehmer の乗算合同法であり(Park and Miller, 1988), ふつうはこれで問題はない(他には加数が素数の線形合同法(L'Ecuyer, 1997)や遅れのあるフィボナッチ数列(Mascagni *et al.*, 1998)など)。

32 ビットで計算を行う場合の「よい」乱数列を与える乗算合同法のアルゴリズムを C 言語でコーディングしたものを以下に示す。これは 32 ビットのオーバーフロー演算を

サポートしていない処理系でも正しく動作する。周期は $2^{31}-1$ になる (Park and Miller, 1988)。

```
double frand()
{
    unsigned long const a=16807L;
    unsigned long const m=2147483647L;
    unsigned long const q=127773L; /* m div a */
    unsigned long const r=2836L; /* m mod a */

    extern unsigned long rnd_seed;

    unsigned long lo, hi;
    signed long test;
    double xf;

    hi=rnd_seed/q; /* / equals to div for int operand */
    lo=rnd_seed%q;
    test=a*lo - r*hi;
    if (test>0) { rnd_seed=test; }
    else { rnd_seed=test+m; }
    xf=(double)rnd_seed/(double)m;
    return xf;
}
```

なお、Knuth 自身が提案している擬似乱数列生成プログラム (<http://www-cs-faculty.stanford.edu/~knuth/programs/rng.c>)も可搬性が高い擬似乱数列発生器とされているが、少々難解である。

シミュレーションの実行

こうして決めた部品を組合せ、全体のプログラムを作るのは難しいことではない。とくに局所変数が見える言語処理系なら他の部分を気にしなくてよいので、組合せ自体は容易である。

その上で直面する問題がいくつかある。まず問題になるのは、初期人口の設定である。安定人口モデルなどに基づいて計算し、合成することももちろんできる(むしろそうすることがふつうである)。しかし、実在する集団のデータを初期人口として与えることを可能にするため、合成するにしても、一旦はファイルに書き出すようにすべきである。また、パラメータだけを変えて結果を比較したいときに、同一の初期人口を使えるようにするためにも、初期人口はファイルから読む方がよい。また、人口規模を何人にするかということも、目的にあわせて慎重に考える必要がある。たとえば、TFR(合計出生率)と避妊の関係を評価するためのシミュレーションを行う場合に、人口規模が 500 人未満では、TFR の指標としての安定性が悪すぎて、避妊との関係をみるという目的に使うには向かなくなってしまう(詳しくは後述)。

次に問題となるのは、何回のシミュレーションを行うべきかということである。これは、もちろん目的による。たとえば、データの生成を目的としてシミュレーションを行う場合は、1回でも十分意味がある。分布を得るためなら、最低でも 1,000 回、ふつうは 10,000 回のシミュレーションが要求される。後述する評価の軸の話とも関係するが、検出力にみあったサンプル数ということである。複雑な事象の再現としての個人レベルの人口再生産シミュレーションの場合なら、10 回とか 100 回でもそれなりに意味がある。最後のケースでは1回のシミュレーションにかかる時間が長いことが多く、10,000 回ものシミュレーションを行うことは非効率的である。回数を決めてしまえば、後は、乱数列の初期値を変えてプログラムを指定回数実行するループを書き、バッチ処理として実行させればよい。

シミュレーション結果の評価

こうして結果は得られるわけだが、まだ大きな問題が残っている。「評価」である。評価の目的は2つある。1つはプログラムの正しさの評価で(プログラムにはバグがつきものであり、それをチェックすることは絶対に必要である)、もう1つはモデルの評価である。

個人ベースの人口再生産シミュレーションでプログラムの正しさの評価することは、意外に難しい。C 言語の場合、プログラムが文法的におかしくないかは lint のようなツールを使えばチェックできるし、cflow のようなツールでプログラム構造を出力させれば、意図した構造になっているかどうかはわかる(注:どちらのツールも unix なら標準であるか port されている)。アルゴリズムのチェックも、個々の関数については既知パラメータを入れて出力結果をみればよい。しかし、それらが正しく組み合わせられたかどうかは、100 人くらいの少人数でシミュレーションを1ないし数ステップやってみて結果を出力させ、目で見ておかしくないかチェックするしかない。動いたと思って安心してはいけないのである。1度目に書いたプログラムでは、想定したメカニズムからは存在するはずのない人がいたり、新生児に ID が割り振られなかったりということがままある。100 人で正しく動いても、それだけでは十分ではない。初期人口を大きくして動かすとメモリアロケーションエラーになってしまうことがある。確保したメモリを解放することを忘れてしまうような場合である。Checker (<http://www.gnu.org/software/checker/checker.html>)

を使えば C のメモリ処理はチェック可能であるため、一応の安定動作が実現できた段階で Checker をかけてみるとよい。

こうして正しく動作するプログラムが得られたとして、次にすべきことはモデルの評価である。つまり、妥当性検証(validation)と感度分析(sensitivity analysis)である。どちらにおいても、結果をどの変数で評価するかということと同時に、表現方法が問題となる。人口再生産シミュレーションで評価対象としてよく選ばれる変数は、人口あるいはその増加率である。目的によってはもっと特異的な変数を選んでもよいが、少なくとも妥当性検証については、実測データが得やすいことが選択基準となる。表現方法としては、平均と標準偏差であらわされることが多いが、注意すべきなのは、一例でも「偶然変わったことが起こりうる」なら重要だということである。つまり、たとえば、同じパラメータで乱数列の初期値だけが異なるシミュレーションを 100 回実行した場合に、100 回中 99 回は人口が単調増加したとしても、残りの1回で集団が絶滅するならば、その1回の試行には大きな意味がある。それゆえ、結果の表現としては、**全部のシミュレーション結果のグラフ表示**を含むべきである。

妥当性検証については、「1度でも実測値と完全に一致するならば、現実がそのシミュレーションで想定したメカニズムの通りに起こった可能性を否定できなくなる」という意味で、完全一致するまでパラメータや乱数を変えて試行を繰り返すのが王道である。Dyke (1981)も MacCluer (1980)も、これを「シミュレーションならではの妥当性検証」として強く支持している。しかし現実には、平均値の変化のグラフを書いて、実測データの変化と重ねて「だいたい合っている」といって済ませている論文が多くみられるのは、完全一致を達成するのが難しいからである(不可測なパラメータを増やせば容易だが、それでは意味がない)。次善の策としては、平均値に標準偏差や標準誤差をつけたり、中央値に上下5パーセンタイルまでの幅をつけ、その範囲内に実測データが入ることを確認するとか、カイ二乗適合度検定で「適合していないとはいえない」ことで消極的に妥当性検証を行うことである。

感度分析とは、モデルが妥当であるといえた上で、個々のパラメータを変えたときの結果の変化を検討することである。「単純な解析モデルが存在しないような複雑な現象のシミュレーション」モデルを作る目的はここにある場合が多い。分布を求めることが目的なら必須ではないが、パラメータの性質を知るためにも有効なので、常に感度分析

はすることを奨める。

個人ベースの人口再生産シミュレーションの実行例

長々とシミュレーション人口学について述べてきたが、最後に実行例を示す。パプアニューギニア低地に居住するギデラという人たちについてのシミュレーションである。これは、個人の定義には配偶関係(一夫多妻婚がある)とマラリア感染、それに対する抵抗性の遺伝子を含み、マラリア流行度の異なる2地域間での移動があるものとし、死亡の発生については Brass のロジットモデルを用い、ペア形成のルールはある年齢差以内のクラン外婚かつ一夫多妻を許すもので(ただし可能なペアについては等確率)、ペアの出生力は扱わず従来通りの女性の出生力にしている、甚だプリミティブなものである。しかし、小集団での抵抗性遺伝子の世代間伝達を考える場合にはペア形成を含むモデルにすることが重要な意味をもつし、出生力のパラメータを検討するためのデータ合成にも役に立つ。初期人口規模を変えて 200 年間のシミュレーションを行って得られる、TFR(合計出生率)と完結パリティの関係プロットしたものを下に示す。人口規模が小さいときの TFR の不安定性が読みとれる(中澤・大塚, 1997)。

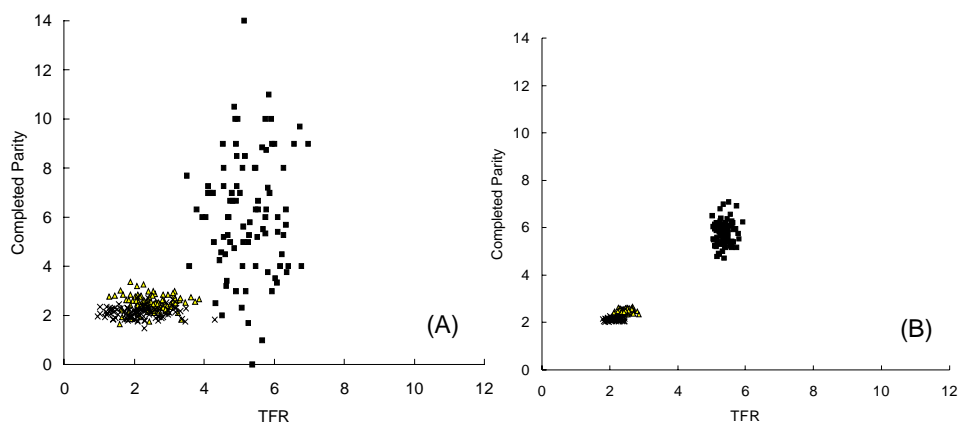


図. TFRと完結パリティの関係(出生・死亡水準: ■「高」; △「中」; ×「低」;
初期人口規模: (A)500人, (B)10000人)

なお、このプログラムは、<http://www.humeco.m.u-tokyo.ac.jp/~minato/demogsim.lzh>としてダウンロード可能なので(プログラムのインストール方法及び使い方については英語マニュアルしか入っていないが)、関心のある方は自分のコンピュータで実行することをお勧めしたい。

謝辞

本稿執筆の機会をお与えくださった巖佐 庸先生ならびに編集部の皆様に深く感謝いたします。また、原稿を読んで助言をくれた研究室の萩原 潤氏にも感謝します。なお、舌足らずな点と冗長な点があつてわかりにくい点は、すべて著者の能力不足からくるものであり、申し訳なく思います。人口学におけるシミュレーションの意義を何か少しでも汲んでいただければ幸いです。

引用文献

- Bongaarts, J. (1978) A framework for analyzing the proximate determinants of fertility. *Pop. Develop. Review*, 4: 105-132.
- Brass, W. (1975) *Methods for Estimating Fertility and Mortality from Limited and Defective Data*. Univ. North Carolina, Chapel Hill.
- Coale, A. J. and P. Demeny (1983) *Regional Model Life Tables and Stable Populations*. 2nd Ed., Academic Press, New York.
- Coale, A. J. and J. Trussell (1974) Model fertility schedules: variations in the age structure of childbearing in human populations. *Pop. Index*, 40: 185-201.
- DeAngelis, D. L. and L. J. Gross, Eds. (1992) *Individual-based Models and Approaches in Ecology: Populations, Communities and Ecosystems*. Chapman & Hall, New York.
- Dyke, B. (1981) Computer simulation in anthropology. *Ann. Rev. Anthropol.*, 10: 193-207. (◆人類学におけるシミュレーション研究のレビュー。古いが、その後も質的に革新的な進歩はないので今日でも価値を失っていない)
- Fredrickson, A. G. (1971) A mathematical theory of age structure in sexual populations: Random mating and monogamous marriage models. *Math. Biosci.*, 10: 117-143.
- Gavrilov, L. A. and N. S. Gavrilova (1991) *The Biology of Life Span: A Quantitative Approach*. Harwood, London.
- Gohel, M., J. J. Diamond, C. V. Chambers (1997) Attitudes toward sexual responsibility and parenting: an exploratory study of young urban males. *Family Planning Perspectives*, 29: 280-283.
- Hu, Y. and N. Goldman (1990) Mortality differentials by marital status: An international comparison. *Demography*, 27: 233-250.
- 稲葉 寿 (1997) 数理人口学の発展. 人口学研究, 21: 7-17.
- カーニハン, B. W., リッチー, D. M. (石田 晴久 訳)(1989) プログラミング言語 C [第2版], 共立出版
- Knuth, D. E. (1981) 「THE ART OF COMPUTER PROGRAMMING 3.準数値算法/乱数」サイエンス社
- 国立社会保障・人口問題研究所 (1998) 第 11 回出生動向基本調査:結婚と出産に

関する全国調査: 夫婦調査の結果概要. 国立社会保障・人口問題研究所, 東京. (◆ <http://www.ipss.go.jp/Japanese/doukou11/doukou11.html>参照)。

- L'Ecuyer, P. (1997) Tables of linear congruential generators of different sizes and good lattice structure. <http://www.iro.umontreal.ca/~lecuyer/myftp/papers/latrules.ps>
- MacCluer, J. W. (1980) The simulation of human fertility - strategies in demographic modeling. In: Mielke, J. H. and M. H. Crawford [eds.] *Current developments in anthropological genetics vol.1 Theory and methods*, Prentice Hall, New York, pp. 111-131.
- MacCluer, J. W., J. V. Neel and C. Napoleon (1971) Demographic structure of a primitive population: A simulation. *Am. J. Phys. Anthropol.*, 35: 193-208.
- Mascagni, M., M. L. Robinson, D. V. Pryor, and S. A. Cuccaro (1998) Parallel pseudorandom number generation using additive lagged-Fibonacci recursions. <ftp://ftp.super.org/pub/mascagni/pprng2.ps.Z> (◆ 発表年度は明示されていない。 <http://141.142.3.76/Apps/CMP/RNG/www/refs.html> から, L'Ecuyer (1997) とともにダウンロードできる文書である。)
- Mode, C. J. and M. A. Salsberg (1993) On the formulation and computer implementation of an age-dependent two-sex demographic model. *Math. Biosci.*, 118: 211-240.
- Nakazawa, M. and R. Ohtsuka (1997) Analysis of completed parity using microsimulation modeling. *Math. Pop. Studies*, 6: 173-186.
- 中澤 港, 大塚柳太郎 (1997) 出生力の指標としての TFR と完結パリティの評価. 人口学研究, 21: 61-63.
- Park S. K. and K. W. Miller (1988) Random number generators: Good ones are hard to find. *Communications of the ACM*, 31(10): 1192-1201.
- Skakkebaek, N. E., E. Rajpert-De Meyts, N. Jorgensen, E. Carlsen, P. M. Petersen, A. Giwercman, A. G. Andersen, T. K. Jensen, A. M. Andersen, and J. Muller (1998) Germ cell cancer and disorders of spermatogenesis: an environmental connection? *APMIS*, 106: 3-11.
- 杉藤重信 (1991) 人口制御要因としての婚姻規制 -コンピュータ・シミュレーションによる分析 - 国立民族学博物館研究報告別冊 15号, 251-275.
- Wood, J. W. (1994) *Dynamics of Human Reproduction: Biology, Biometry, Demography*. Aldyne de Gruyter, New York. (◆ ヒト出生力研究の最高の Text. 653 ページ一杯に詰め込まれた情報が約 5,000 円で入手できるのは素晴らしい。ただし男性の寄与については最終章で短く触れているだけなのが惜しい)
- Wood, J. W., D. J. Holman, K. M. Weiss, A. V. Buchanan, and B. LeFor (1992) Hazards models for human population biology. *Yearbook Phys. Anthropol.*, 35: 43-87. (◆ 生物人口学におけるハザード解析のレビュー。実例が豊富でわかりやすい)