TASK QUARTERLY 7 No 2 (2003), 263–269

# POPULATION STRUCTURE IN THE PENNA MODEL

MIKOŁAJ SITARZ

*Department of Physics and Nuclear Techniques,*
*University of Mining and Metallurgy,*
*Mickiewicza 30, 30-059 Cracow, Poland*
*mik@fatcat.ftj.agh.edu.pl*

(Received 27 July 2002; revised manuscript received 23 January 2003)

**Abstract:** Numerical simulations of growing population, based on the Penna model, are presented. We consider biological ageing as a result of the accumulation of mutations in the genotype. We present an example of simulated population structure and compare it with that of a real population. We also discuss the role of bad mutations threshold parameter $T$ in the Penna model, to reproduce a smooth transition from the oversimplified logistic model ($T \to \infty$) to the Penna model.

**Keywords:** biological ageing, Penna model, bit-string model, population structure

## 1. Introduction

The logistic model considers two factors: the birth rate and the probability of death due to limited environmental capacity. In a simulation, death happens with the probability known as the Verhulst factor: $n/N$, where $n$ is the number of individuals at a given time, and $N$ is the environmental capacity. This would lead to a new population at time $(t+1)$:

$$n(t+1) = n(t) - \frac{n(t)}{N} n(t) = n(t)\left(1 - \frac{n(t)}{N}\right), \tag{1}$$

where $n(t)$ is the population at time $t$. However, we need to consider another step in the simulation, when the survivals give birth to babies. Then, with the birth rate $B$, we get:

$$n(t+1) = n(t)\left(1 - \frac{n(t)}{N}\right) + B \cdot n(t)\left(1 - \frac{n(t)}{N}\right), \tag{2}$$

or

$$x(t+1) = k \cdot x(t)(1 - x(t)), \tag{3}$$

where $x(t) = n/N$, and $k = B+1$. This is the well-known logistic equation.

Therefore, each evolution cycle consist of the following two consecutive steps:

- decrease of the population according to the current $n/N$ ratio, the so-called Verhulst factor;

- increase of the remaining population by the fraction $B$ (that means, that one individual gives $\lfloor B \rfloor$ descendants, where $\lfloor B \rfloor$ is an integer part of $B$, and an extra one with probability $P(B) = B - \lfloor B \rfloor$).

Depending on the $B$ parameter, the solution (population in a stationary state) is stable, periodic or chaotic. The logistic equation predicts a stable solution $x^* = B/(B+1)$ for $B$ between 0 and about 2, a periodic for $2 < B < 2.6$, and a chaotic one above 2.6. Since analytical solutions are known, this simple model is useful for reference and testing Penna model [1] simulations (when it is reduced to the logistic case). The Penna model may be seen as a generalization of the perhaps oversimplified logistic model. In this model, each individual has an inherited bit-string genome (the entire genome is an integer). In each bit position, bit '1' is interpreted as the presence of a bad mutation (a disease), and bit '0' – as the lack of it. Bits are disclosed according to the actual age of an individual: if its age is $a$, the first $a$ bits are checked for number $\mu(a)$ of mutations. These mutations are considered active. So, $\mu(a)$ is pre-determined at birth.

The Penna model incorporates a new elimination mechanism into the logistic model, according to an algorithm enriched by the possibility of a genetic death occuring when too many active mutations $\mu(a)$ are active. The simulation steps are then as follows:

- possible death caused by the Verhulst factor,
- genetic death if number of active mutations $\mu(a)$ reaches a threshold value $T$, then
- birth of $B$ babies of mature items with minimum reproduction age $R$. The offspring inherits its genome from its parent (an asexual model), enriched by $M$ additional mutations randomly picked in the whole bit-string.

Thus the standard Penna model has 5 input parameters $(N, B, R, M, T)$. As a result of the simulation, the population $n(t)$ at a time step $t$ may now be more precisely characterized by some more detailed data, such as $n(t,a,\mu)$, the population fraction of a given age $a$ and of active mutations $\mu$. Then, the total population at time $t$ is given as:

$$n(t) = \sum_{a,\mu} n(t,a,\mu). \tag{4}$$

Since we are interested in a stationary population, we consider $n$ after a long time:

$$n(a,\mu) = n(t \to \infty, a, \mu), \tag{5}$$

and then the population group at age $a$ is

$$n(a) = \sum_{\mu} n(a,\mu). \tag{6}$$

We may define another parameter useful in the analysis of results, the survival rate:

$$s(a) = \frac{n(a+1)}{n(a)}, \tag{7}$$

which is the fraction of individuals survived to another evolution cycle. We may also use the mortality rate defined as:

$$q(a) = 1 - s(a). \tag{8}$$

The Penna model in the limit of the input parameters set $(N,B,R,M,T) = (N,B,0,0,\infty)$ is reduced to the logistic case. The logistic limit of the Penna model may be used as a test for simulations, since analytical results are then known. For the logistic limit of the Penna model for a stationary state, we get:

$$q(a) = \frac{B}{1+B}, \tag{9}$$

which is $a$-independent, and we can similarly obtain:

$$\frac{n(a)}{N} = q^2 s^a, \tag{10}$$

from which the total population is

$$\frac{n}{N} = \frac{\sum n(a)}{N} = \frac{B}{1+B}. \tag{11}$$

The Penna model can easily be enriched by various modifications, as discussed in [2, 3], which may essentially broaden the scope of the model.

## 2. Discussion of results for standard Penna model

In this section we compare the Penna model simulation results, with some data on the population in Poland in the years 1993–1997 (source [4]).

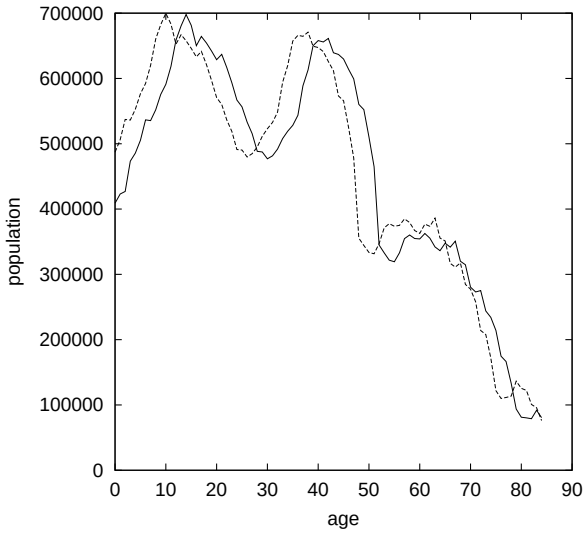


**Figure 1.** Age structure of the population in Poland; the solid line is for year 1997, dashed – 1993

The age structure according to the statistical data shown in Figure 1 is far from the standard model results predicted by the Penna model. The Penna model predicts a monotonic decrease in $n(a)$ dependence. However, the mortality coefficient $q(a)$ is a quantity which is insensitive to the actual population size and, therefore, the sudden drops or peaks in the $n(a)$ plot may disappear for the $q(a)$ plot.

This strong deviation originates from the fact that the model parameters, such as birth rate $B$, may be time dependent, for example reflecting the current tendency

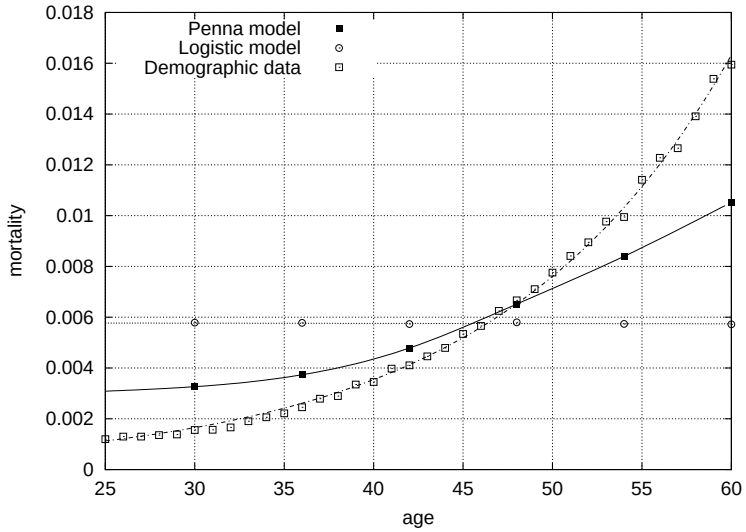


**Figure 2.** Mortality rate (arbitrary units) for the population in Poland (1996-1997), with the Penna and logistic model simulation results. The solid line corresponds to the Penna model with splines, the dotted (linear) – to the logistic model, the dashed (exponential) – to demographic data

towards families with less children. The discovery of a new medicine may also upset the death rate. Or migration out of a country in difficult times may eliminate individuals at the most active age. Perhaps a dip in the age structure around 50 is a reminiscence of the War times?

We conclude that only a representative group must initially be chosen for an analysis of age distribution in a population, perhaps from among the more stable countries. Mortality rate:

$$q(a,t) = 1 - \frac{n(a+1,t+1)}{n(a,t)}, \tag{12}$$

however, is expected to be less sensitive to the population fluctuations versus age. This is so since only the ratio $n(a+1,t+1)/n(a,t)$, relevant to the formula for mortality, is not sensitive to the population $n(a)$ itself.

The actual size of a population should not influence significantly the mechanism creating the balance between the birth rate and the genetic death toll, which is responsible for mortality. Figure 2 shows the mortality $q(a)$ distribution based on the same data as in Figure 1, marked with empty squares. A simulation (full squares) for the $(N,B,R,M,T) = (10^7,1.2,4,1,4)$ set of parameters has been carried out on a 32-bit machine. The computer time-step must be converted to a physical time scale so that the maximum age of real population and the simulation of the oldest population members coincide. On the computer, it is about $a = 16$ which corresponds to the maximum age 90, and so a factor of $k = 6$ has been used as a conversion factor between computer age and physical age. We have considered only the middle range of ages, where the Penna model results have its exponential region. In this case, it corresponds to the range $a \in (25,60)$ in demographic data.

It can be seen that the agreement is reasonable and the main $q(T)$ tendency is largely fulfilled, but there are some difficulties in applying the Penna model

simulation results to demographic data. These days, birth rate $B$ is quite low for human populations in a lot of countries (see [5, 6]). It amounts to mere $0.02-0.03$ (average birth rate for one individual per year). Unfortunately, the Penna model cannot be applied for such values, because it leads to a zero-size population (see Figure 3). Changing the $R$ parameter, even to zero, doesn't change this significantly.

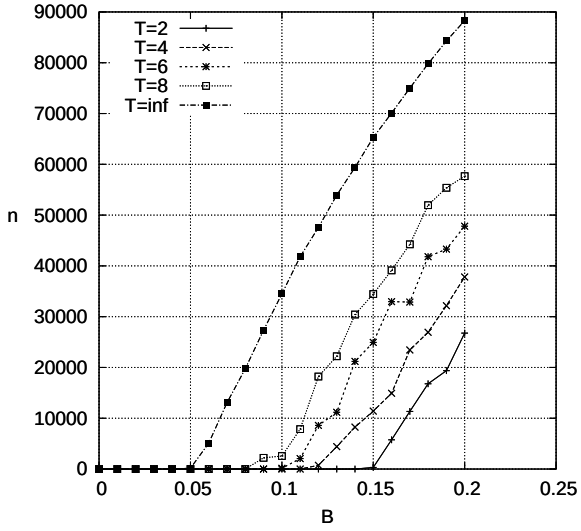


**Figure 3.** Population $n$ in a stationary state as a function of birth rate $B$ for various $T$. Simulation for the standard Penna model $(N,B,R,M,T)=(10^6,(0;0.2),4,1,\{2,4,6,8,\infty\})$

As can be seen, the upper limit of the $B$ parameter, for which we get $n=0$, depends on $T$. The critical $B$ is the least for $T\to\infty$ (the logistic case) and is a decreasing function of it. The small (non-zero) critical $B$, even for the logistic model, results from the fact that the true logistic model does not have an upper limit of individual age, as in our simulation (where it equals the length of the genome).

Some other interesting results, including a comparison of demographic data and simulated mortality rate, are to be found in [7].

## 3. Influence of parameter $T$ on the mortality rate

The threshold $T$ plays an essential role in the Penna model, being directly responsible for the occurrence of genetic death. Therefore, the $T\to\infty$ limit may be seen as the logistic model limit when only limited environmental capacity $N$ eliminates items from the population. So, we intend to scan various values of $T$ in the simulation, and the results may be discussed in terms of the Gompertz law. This empirical law states that mortality $q(a)$ increases exponentially with age $a$:

$$q(a)=q_0\cdot e^{ba}. \tag{13}$$

However, in the logistic model limit $T\to\infty$ we get $q(a)=B/(B+1)$, and so $b=0$. Therefore, it may be interesting to observe how the Gompertz law parameters are modified when we change $T$:

$$q(a)\to q(a,T)=q_0(T)e^{b(T)a}. \tag{14}$$

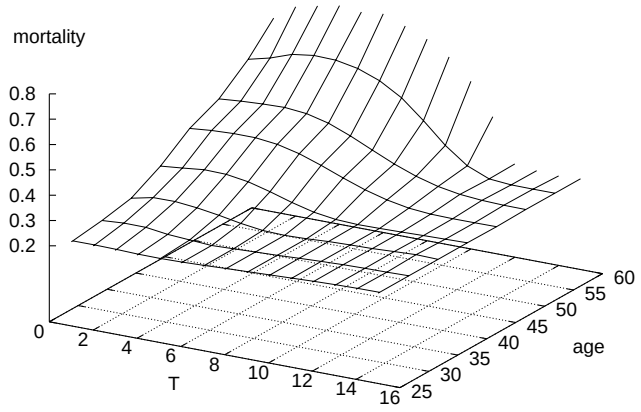


**Figure 4.** Mortality rate for different $T$ parameters; the standard Penna model $(N,B,R,M,T) = (10^7,1.2,4,1,(1;15))$

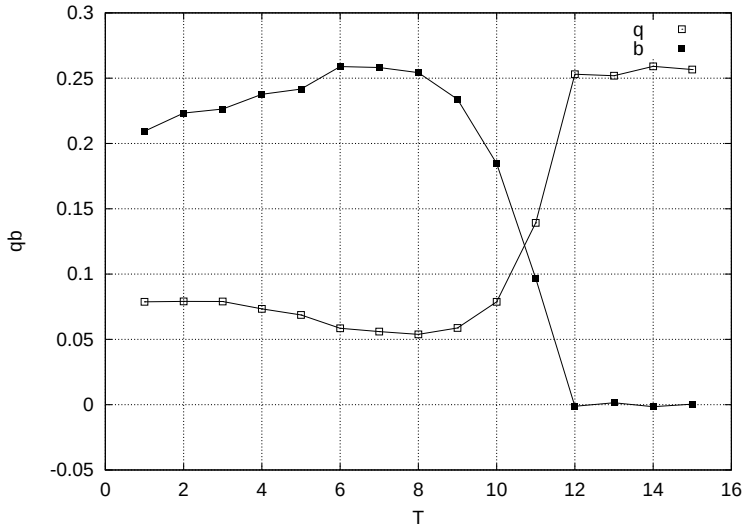


**Figure 5.** $q_0$ and $b$ as functions of $T$

Figure 4 presents how the mortality rate depends on age for various values of $T$ ($T \in (1;15)$). It has been done for the same region which has been used to produce Figure 2. As has been expected, $q(a)$ is becoming flatter and flatter as we go towards the highest $T$. An even better view of this is given in Figure 5. It demonstrates how the parameters $q_0$ and $b$ (from Equation (13)) depend on $T$. To obtain these dependencies, we have used the same results as in the previous chart. Then,

- for each $T$ (15 cases), we fitted a curve (13), using only values for $a \in (25,60)$. The curves were fitted using the nonlinear least-squares Marquardt-Levenberg algorithm;

- from each fit, we got a pair $(q_0(T), b_0(T))$;
- both dependencies were put on the chart.

This result may be used to predict the influence of $T$ on $q(a)$. As can be seen, for $T$ from 1 to about 8, the $q_0(T)$ and $b(T)$ dependencies are monotonic. They have their extrema at $T \approx 8$, followed by another monotonic section. A sharp transition at $T = 12$ is caused by the fact that we have considered a limited range of ages. For that particular $T$, when the flat region is large enough to be wholly included in our range, we get an age-independent mortality rate

$$q(T) = q_0 \cdot e^0 = q_0, \tag{15}$$

and an increasing $T$ no longer influences this value. This simulation has been done for one set of parameters only and we do not know how it will change when $R$ or $M$ are altered, for example. However, our partial results (for other sets) suggests, that the tendency shown in the figure will not change significantly. More about these factors can be found in [8].

## 4. Conclusions

We conclude that it should be possible, to use the Penna model for reconstructing the mortality rate of a real population. Perhaps a more precise agreement between the data and the results of the Penna model may be achieved if we try to match more parameters of the model, not only the $k$ factor. Results that reproduce the transition between the logistic and the Penna model are satisfactory.

### *Acknowledgements*

I would like to thank Prof. Andrzej Maksymowicz for all his suggestions. Computer simulations have been done at the CYFRONET Computer Center, Cracow, Poland, on an SGI 2800 machine.

### *References*

[1] Penna T J P 1995 *J. Stat. Phys.* **78** 1629
[2] Magdoń M S and Maksymowicz A Z 1999 *Physica* **A 273** 182
[3] Maksymowicz A Z 1999 *Physica* **A 273** 150
[4] GUS Poland, http://www.stat.gov.pl/
[5] Population Reference Bureau, http://www.prb.org/
[6] U.S. Census Bureau, http://www.census.gov/
[7] Racco A, Argollo de Menezes M and Penna T J P 1998 *Theory in Bioscience* **117** (2) 101
[8] de Oliveira S M, de Oliveira P M C and Stauffer D 1999 *Evolution, Money, War and Computers*, Teubner, Leipzig-Stuttgart