Annales UMCS Informatica AI 3 (2005) 205-211

Annales UMCS
Informatica
Lublin-Polonia
Sectio AI

http://www.annales.umcs.lublin.pl/

Galton Board with memory

Jerzy L.Kacperski*

Computer Laboratory 2, Institute of Physics, University of Łódź,
Pomorska 151, 91-236 Łódź, Poland

Abstract

The Galton Board model complemented with a simple memory mechanism was employed to illustrate the Galton Table structure

1. Introduction

Real and tangible equivalents of the items so loved by statisticians and used to carry out mental experiments – coins, dice, a deck of cards or an urn of balls, would in reality be affected by the destructive forces of physical surroundings, and, consequently, their properties would change. Additionally, when we speak of an *ideal* coin or dice, we implicitly mean not only their unchangeability but also exact symmetry, as a consequence of which, all elementary events are equally probable. It's worth remembering though, that a two-state symmetrical coin would have to be infinitely thin and have no heads or tails; similarly an ideal dice would have to have identical sides. As a result, the indistinguishability of final states would in each case be a not-necessarily desired consequence of exact symmetry.

In contrast to real coins and dice, their virtual equivalents can be unchangeable, but they cannot be ideal. This is because, programmatic generation of numbers other than *pseudorandom* is impossible. *"... Anyone who considers arithmetical methods of producing random digits is, of course, in a state of sin ..."* [1]. Building a good hardware white noise generator is also a difficult task.

2. The Galton Board

Among statistical equipment encountered not so long ago in introductory physics labs, the Galton Board stands out. It is a device, "…*which wonderfully*

**E-mail address*: jlk@uni.lodz.pl

illustrates a whole class of physical phenomena, which we tend to call diffusion and heat transfer… [2].

The Galton Board, aka quincunx board (because of its regular lattice structure with a square/rectangular centered cell) is a vertically sloping or vertical board with n layers of rods (scatterers) indexed with $0,1,\ldots,n-1$ and an input slot, through which identical balls come in, one by one, or at least serially, thanks to which, they move independently. Each ball scatters in each layer, each time deflecting horizontally at a value equal to a where $2a$ is the lattice constant, i.e. the ball moves along the branches of the binary tree (let's say the deflection to the right occurs with a probability of p) and eventually, after n collisions, falls into *k-th* ($k=0,1,\ldots,n$) bin at the bottom of the table, where k is the number of deflections 'on the right'. For simplicity reasons, we put the lattice constant (and also the bin width) $2a=2$. The balls gathering in the bins create gradually growing histogram bars, empirical distribution resembling the corresponding binomial distribution $B(n,p;k)$. In some sense, the Galton Board is an instance of the Pascal Triangle class. If the bin borders are located below the rods of the lowest layer, then the numbers of possible paths to each rod (binomial coefficients) are known, and obviously these numbers in the n-th (fictitious) layer will be proportional to the expected bins population. The mean, variance and average squared value of the random variable k are $n/2$, $n/4$, and $(n^2+n)/4$, respectively. After introducing the new statistics, $x=2k-n$, the final distance from the projection of the input slot on the table basis can be for example interpreted as caused by how random errors influence the deviation of the measurement result from the 'true' value symbolized by the location of the slot μ, or the expected value of normally distributed random variable $N(\mu,\sigma^2)$ where $\sigma^2=n$; i.e. the histogram approaches the normal distribution when the number of layers $n>>1$.

The bell-shaped histogram of the final deflections is usually treated as a metaphorical illustration of the CLT (Central Limit Theorem) – an approximation of the normal distribution for any independent trials process such that the individual trials have finite variance as in the case of the above-mentioned Bernoulli trials process.

The ball wandering through the table is the same as the n-fold coin toss and the title of the 3rd Chapter of the classic book [4]: *Fluctuations in Coin Tossing and Random Walks* could be undoubtedly replaced with: *Galton Board and Random Walks*; and, similarly, in a very often cited opinion from the very same work, about the significance of the experimental (in practice computer-aided) approach in statistics and probability theory: *„The results (…concerning fluctuations …) are so amazing and so at variance with common intuition that even sophisticated colleagues doubted that coins actually misbehave as theory predicts. The record of a simulated experiment is therefore included*."

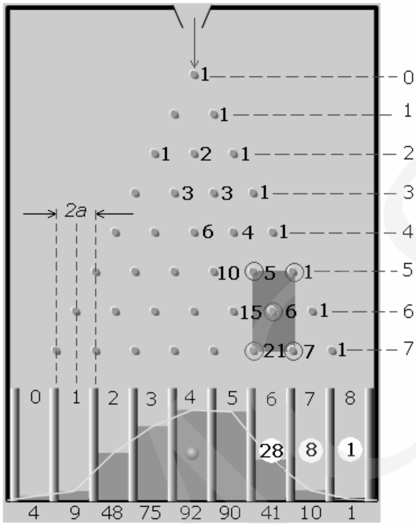


Fig. 1. A screenshot of the Flash movie frame [5] with a scheme of indexing layers and bins. The bolded figures denote the number of possible paths to the given rod (scatterer)

Apart from formal equivalence, the coin-aided experiment is no match for self-explanatory Galton Board. When for example the Brownian motion is observed in a student laboratory, with particle paths projected onto the microscope focal plane and only one motion component measured [4], the similarity of a Brownian particle to the steel ball wandering across the table, is striking – both these processes represent a random walk in one direction. The role of time in the Einstein's relation $<x^2>\sim t$ for the average squared distance of the particle from the diffusion center is played by, in the case of the board, the number of scattering layers n.

3. Implementation

The importance of simulated chance experiments in explaining and interpreting statistical results increases with the overall role of computers in the experiment. When the implementation of virtual border [5] was chosen as an element of virtual laboratory in probability and statistics, the preferences of Internet users (short file download time) were taken into account. According to the results of the analysis [6] of some of the advantages of the most widely used technologies from the point of view of IP&A (Image Processing and Analysis) Flash movie is about 2.4 smaller and its download time is more than 2.6 shorter in comparison with a still very popular Java applet. The Galton board implementations using both applets (e.g. [7]) and Flash movies (e.g. in a virtual laboratory [8]) were published on the Internet.

4. Galton Board with memory

The real Galton Board cannot be too good an approximation of the exemplary device, where the right or left shift should occur with the same probability (with default being *0.5*). In a mechanical board the scatterers are biased, i.e. the parameters of distributions slightly differ and, additionally, the consecutive deflections are probably correlated, which could be called *a memory* – not ball memory, because the ball is devoid of internal structure, but board memory.

Francis Galton, the inventor of quincunx is also known (probably even better) as the author of probability models explaining the genetic effect on a person's height [9]. The analysis of the correlation (in modern terminology) between children's (when they are adult) and their parents' heights gave birth to one of the most important ideas in statistics – regression to the mean. The empirical distribution of height was normal with great accuracy and this suggests this attribute was influenced by many small independent factors; but it was well known that height is sufficiently determined by the genetic factor. With the help of a simple procedure modeling Galton board with memory (see Appendix 1) we collected a sample of size of the original (928 children and 207 parents) and obtained the table presented in Appendix 2. Joint density function *fPC(p,c)* of the random variable *parent-child* (*PC*) is presented in Fig. 1. After smoothing (not shown) the constant density points form the ellipse $p^2\text{-}2\rho pc+c^2=const$, where ρ is the coefficient of correlation.

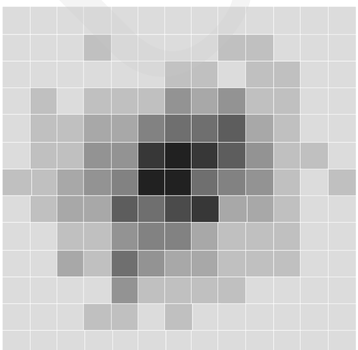
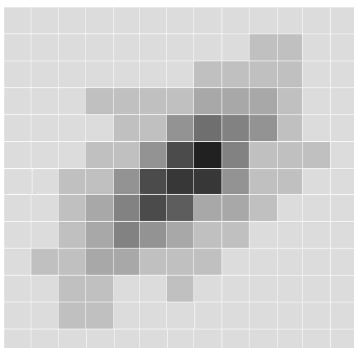
a)  *b)* 

Fig. 2. The characteristic features of Galton table reproduced by Galton board with memory for *pP=0.5, pC=0.6 (a)* and *pP=0.5, pC=0.8 (b)*

The relationship between the Galton board and the Galton table is evident only when memory is put into the board, thanks to which the descendant may 'inherit' an inclination to certain deviations. It is non-biological inheritance however – in a computer table you can invert the time and check that when fathers and sons swap places, the result remains unaffected (not mentioning the fact that regression to mean is most evident when there is no correlation).

5. Conclusions

The ‘illustrative power’ of the Galton Board, an excellent statistical device, rises sufficiently after including indeterministic (fuzzy) memory into model; fuzzy means here that the saved present state of the *parent* ball in a given layer modifies future *child* ball behavior only with some probability $p\leq 1$. This simple approach would allow us to visualize such important statistical concepts as correlation and regression (in both senses – Galton’s [10] and modern).

A virtual table with memory could function in a student biology lab, just as its special case – the ‘amnesia board’ – functions in a physics lab. Modeling could illustrate not only the ancestor-descendant relationship just as in Galton’s experiment where the diameters of subsequent generations of sweet pea seeds were investigated, but also the results of repeated tests on the same object, when regression toward mean might suggest that certain systematic effects exist, such as better/worse test results as a consequence of practice/fatigue (a typical example of regression fallacy).

Appendix 1

```
/* the two-dimensional array ile[][] is filled on the
assumption that the saved track of the parent ball modifies
the probability of deflection of the child ball (from pP to
pC); after pP=pC the memory is cleared*/
function deflection(m)
{
  var neg=m==0?1:0;
  p=m==-1?pP:pC;
  return Math.random()<=p?Math.abs(m):neg;
}
function ft(n)
{
  var n0,ch='0';
  for(n0=1000;n0>1;n0/=10)
  ch+=n<n0?'0':'';
  document.write(ch+n+'\u0020');
}
for(i=0;i<=l;i++)
for(j=0;j<=l;j++)ile[i][j]=0;
for(t=1;t<=tm;t++)
{
if(t%2==1&&t%4==1)
  {
  i=0;
  for(k=0;k<l;k++)
    {
    d[k]=deflection(-1);
    i+=d[k];
```

```
      }
    i0=i;
    }
else
    {
    j=0;
    for(k=0;k<l;k++)j+=deflection(d[k]);
    j0=j;
    }
if(t%2==0)ile[i0][j0]+=1;
}
for(j=l;j>=0;j--)
{
for(i=0;i<=l;i++)
ft(ile[i][j]);
document.write('<br />');
}
```

Appendix 2

Table 1. The imitation of Galton cross-tabulation of 928 adult children by their height and their parents' height

		parent ball deflection													totals	medians
		0.5	1.5	2.5	3.5	4.5	5.5	6.5	7.5	8.5	9.5	10.5	11.5	12.5		
child ball deflection	12.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-
	11.5	0	0	10	1	0	0	0	0	1	1	0	0	0	10	2.2
	10.5	0	0	10	0	0	2	1	3	0	2	2	0	0	10	8.0
	9.5	0	1	49	1	3	8	11	9	10	3	3	0	0	49	7.1
	8.5	0	1	110	3	7	22	24	28	12	7	3	0	0	110	6.7
	7.5	0	1	196	8	14	40	46	40	29	10	4	1	0	196	6.7
	6.5	1	1	205	12	18	45	49	26	32	11	2	0	1	205	6.4
	5.5	0	1	179	12	28	29	38	42	16	6	2	0	0	179	6.2
	4.5	0	3	105	9	21	20	24	13	8	3	3	0	0	105	6.0
	3.5	0	0	60	4	10	17	7	12	2	2	1	0	0	60	5.8
	2.5	0	0	8	0	2	2	1	1	2	0	0	0	0	8	6.4
	1.5	0	0	3	1	0	0	1	0	1	0	0	0	0	3	6.2
	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-
totals		1	8	24	51	103	185	202	174	113	45	20	1	1		
medians		6.5	6.4	6.0	6.0	5.8	6.4	6.6	6.7	6.8	7.1	7.3	7.5	6.5		

References

[1] von Neumann J., *Various Techniques Used in Connection with Random Digits*, NBS Appl. Math. Ser., Washington, DC: U.S. Government Printing Office, 12 (1951) 36.
[2] Smoluchowski M., *Selected Philosophical Papers*, Warsaw, (1956) 286 (in Polish).
[3] Feller W., *An Introduction to Probability Theory and Its Applications*, John Wiley&Sons, Inc., 1 (1981).
[4] Kacperski J.L.,Sonarski T., *Advanced Physics Laboratory*, Lodz University, (1986) 141 (in Polish).
[5] Kacperski J.L., Przybył W., *Galton Board*, http://fizjlk.fic.uni.lodz.pl/gb
[6] Zhu X.Q., Zhang Y.J., Liu W.J., *Evaluation and Comparison of Web Course Developmental Tools and Technology: A Case Study*, http://www.stanford.edu/~zhuxq/publication/ICCE_compare.pdf
[7] Pluta J., *The Galton Board or the beginning of an Internet Physics Laboratory?!,* http://www.if.pw.edu.pl/~pluta/pl/galton-pl.html (in Polish)
[8] *Interactive Mathematics Projects using Macromedia Flash*, http://teacherlink.org/content/math/interactive/flash/
[9] Grinstead C.M., Snell J.L., *Introduction to probability*, American Mathematical Society, 519.2 G868 (1997) 345 and Appendix B, http://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/book.html
[10] *It [regression] is a universal rule that the unknown kinsman in any degree of any specified man is probably more mediocre than he (Francis Galton, 1886), retrieved from:* http://www.tc.umn.edu/~wilson/bae8013/download/simple.pdf