Extending Galton's Binomial Quincunx to the Trinomial Septcunx

Jennifer Harlow*   Bry Ashman†

Raazesh Sainudiin‡

An Article Submitted to

Technology Innovations in Statistics Education

Manuscript 1026

*Department of Mathematics and Statistics, Private Bag 4800, University of Canterbury, Christchurch, New Zealand,

†Department of Mathematics and Statistics, Private Bag 4800, University of Canterbury, Christchurch, New Zealand,

‡Laboratory for Mathematical Statistical Experiments and Department of Mathematics and Statistics, Private Bag 4800, University of Canterbury, Christchurch, New Zealand, r.sainudiin@math.canterbury.ac.nz

Abstract

This paper discusses the development of graphical user interfaces (GUIs) to illustrate sampling from a trinomial distribution by the natural extension of Galton's Quincunx to three dimensions.

1. Introduction and Motivation

The visual or graphical display of data as part of statistical analysis is commonly dated to the late 18th century when Priestly and Playfair used graphics to illustrate their works. Since Tukey's influential *Exploratory Data Analysis* (Tukey, 1977), visual displays have become increasingly important. Developments in computing power and graphics hardware and software have been exploited to be able to handle and visualise datasets of a size and complexity that would probably stun Tukey, let alone Playfair (Card, Mackinlay & Shneiderman, 1999). The principles for good visual displays, such as those set out in Tufte's classic *The Visual Display of Quantitative Information* (Tufte, 1983) and the deployment of computing power for the graphical display of statistical data have been focused on rapid data analysis and interpretation.

For students, graphics may need to fulfil a different role. When data are displayed, often it is the illustration of concepts through data rather than analysis of the data itself that is important; the emphasis changes from information density to information clarity. Computer packages offer powerful graphics capabilities for data analysis but learning to use the package may distract from the central point of understanding the concepts (Nolan & Temple, 2003). Sedig & Liang (2008) use the term visual cognitive tools (VCTs) for the "external mental aids that maintain and display visual representations of information", where 'information' here means "structures, objects, concepts, ideas and problems" (Sedig & Liang, 2008, p. 147), not just the data in an investigation. Graphical displays which are ideal for data analysis have to be adapted for students in order to become an effective VCT to illustrate the actual processes undertaken by the statistician or elucidate an underlying principle.

This paper presents a project to create a visual cognitive tool (VCT) for graphically illustrating the construction of the binomial and trinomial random vectors from the sum of independent and identically distributed de Moivre random vectors in two and three dimensions. We would like the students to understand how arithmetic in the real vector space (two and three dimensions, in particular) allows for the constructive definition of multinomial random vectors (binomial and trinomial, in particular) from the sum of ortho-normal basis vectors chosen with equal probabilities – the so called de Moivre random vectors. We would like the students to visually appreciate a set of realisations of the binomial and trinomial random vector, the corresponding population and sample means, empirical sample frequencies, probability mass function and the geometric meaning of the binomial and trinomial coefficients. Our VCT achieves these learning objectives by allowing the student to interact with graphical illustrations that formalise and extend Sir Francis Galton's ideas from over 120 years ago. Without our VCT only the top 2% of the students are able to appreciate the construction by visualising on their own from the mathematical description of the construction given in Models 1, 2, 3 and 4 of § 2.

The VCT is used in a second year Computational Statistics course in a three year B.Sc. degree program at the University of Canterbury, New Zealand. This is comparable to a junior or senior level course in a four year liberal arts Baccalaureate degree program with a major in Mathematics and/or Statistics in the United States of America. Students are expected to have already taken a 100 level Mathematics, Statistics, or Computer Science course. Students use the VCT in the weekly laboratory sessions to have the constructively defined binomial and trinomial random vectors that were covered during the two weekly lectures reinforced through visual interactions. All necessary mathematical concepts for the appreciation of the binomial and trinomial distributions in this project, including integer vector arithmetic in two and three dimensions and the basic probability models are introduced earlier in the course from first principles. Students of this "YouTube generation" generally respond well when the mathematical, statistical and computational concepts that are traditionally communicated in lectures using the predominant read/write style of teaching/learning are also concurrently communicated in an interactive manner via visual cognitive tools in the weekly laboratories.

We start by revisiting Sir Francis Galton's Quincunx. We formalise a probability model of the Qunicunx as a binomial random vector that is obtained as the sum of independent and identically (IID) distributed two-dimensional Bernoulli random vectors. We show how the Quincunx can be used to demonstrate the binomial coefficient graphically with the paths leading to an outcome. We extend the Quincunx in a natural manner to its three-dimensional version called the Septcunx. The probability model for the Septcunx is the trinomial random vector obtained as the sum of IID three-dimensional random vectors. Again, the trinomial coefficient can be demonstrated graphically.

2. Sir Francis Galton's Quincunx

We first quote the following words from Sir Francis Galton's historic work, *Natural Inheritance*, pp. 62-65, Macmillan, 1889. Here you will find the kernels behind the construction of our Model 2, the Binomial(n, θ_1, θ_2) random vector $(\vec{\mathrm{RV}})$ as the sum of n IID Bernoulli(θ_1, θ_2) random vectors $(\vec{\mathrm{RV}}\mathrm{s})$ of our Model 1, Bernoulli's Weak Law of Large Numbers and the Central Limit Theorem.

"Mechanical Illustration of the Cause of the Curve of Frequency. – The Curve of Frequency, and that of Distribution, are convertible : therefore if the genesis of either of them can be made clear, that of the other also becomes intelligible. I shall now illustrate the origin of the Curve of Frequency, by means of an apparatus shown in Fig. 7, that mimics in a very pretty way the conditions on which Deviation depends. It is a frame glazed in front, leaving a depth of about a quarter of an inch behind the glass. Strips are placed in the upper part to act as a funnel. Below the outlet of the funnel stand a succession of rows of pins stuck squarely into the backboard, and below these again are a series of vertical compartments. A charge of small shot is inclosed. When the frame is held topsy-turvy, all the shot runs to the upper end; then, when it is turned back into its working position, the desired action commences. Lateral strips, shown in the diagram, have the effect of directing all the shot that had collected at the upper end of the frame to run into the wide mouth of the funnel. The shot passes through the funnel and issuing from its narrow end, scampers deviously down through the pins in a curious and interesting way; each of them darting a step to the right or left, as the case may be, every time

Figure 1: Figures from Sir Francis Galton, F.R.S., *Natural Inheritance*, Macmillan, 1889.
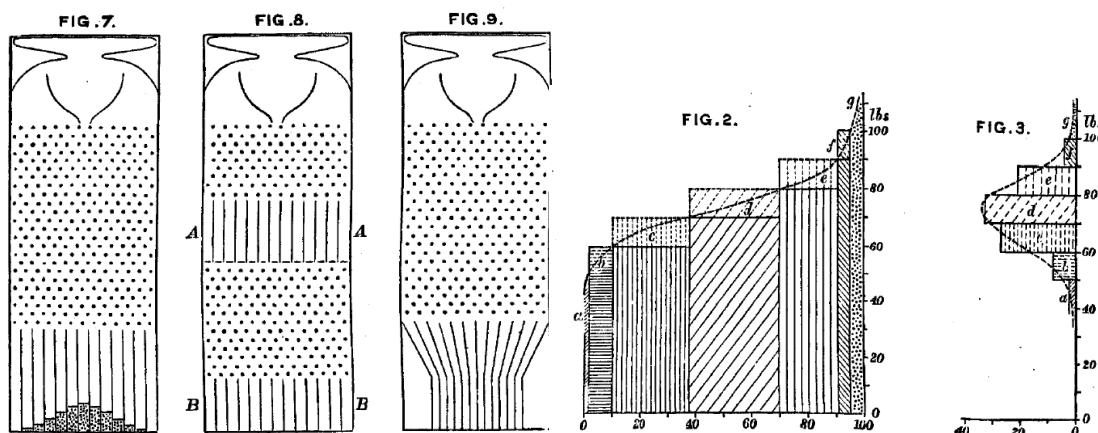
(a) FIG. 7, FIG. 8, and FIG. 9 (p. 63)   (b) FIG. 2 and FIG. 3 (p. 38)

it strikes a pin. The pins are disposed in a Quincunx fashion, so that every descending shot strikes against a pin in each successive row. The cascade issuing from the funnel broadens as it descends, and, at length, every shot finds itself caught in a compartment immediately after freeing itself from the last row of pins. The outline of the columns of shot that accumulate in the successive compartments approximates to the Curve of Frequency (Fig. 3, p. 38), and is closely of the same shape however often the experiment is repeated. The outline of the columns would become more nearly identical with the Normal Curve of Frequency, if the rows of pins were much more numerous, the shot smaller, and the compartments narrower; also if a larger quantity of shot was used.

The principle on which the action of the apparatus depends is, that a number of small and independent accidents befall each shot in its career. In rare cases, a long run of luck continues to favour the course of a particular shot towards either outside place, but in the large majority of instances the number of accidents that cause Deviation to the right, balance in a greater or less degree those that cause Deviation to the left. Therefore most of the shot finds its way into the compartments that are situated near to a perpendicular line drawn from the outlet of the funnel, and the Frequency with which shots stray to different distances to the right or left of that line diminishes in a much faster ratio than those distances increase. This illustrates and explains the reason why mediocrity is so common."

We will mathematically present these concepts in the sequel as a way of giving precise meanings to Galton's observations with his Quincunx and its three-dimensional extension to our Septcunx. Then we show how our GUI interactively builds the binomial and trinomial distributions from these historic devices. In the process we will animate how the sum of bivariate Bernoulli and trivariate de Moivre random vectors yield the binomial and trinomial random vectors, respectively.

We can relate a discrete probability model to the Quincunx. First, we need to introduce simple random vectors ($R\vec{V}$), i.e. ordered pairs, ordered triples, or more generally ordered m-tuples of random variables $(X_1, X_2, \ldots, X_m)$. We focus on elementary definitions needed

to define bivariate R$\vec{\text{V}}$ obtained from a pair of random variables. Here is a simple example of a discrete bivariate R$\vec{\text{V}}$ that illustrates the notions of joint and marginal probabilities.

Example 1. *Recall that a fair coin toss with the outcome* `Heads` *encoded as* 1 *and the outcome* `Tails` *encoded as* 0 *is modelled by the* Bernoulli(1/2) *random variable (RV). Let X_1 and X_2 be a pair of IID* Bernoulli(1/2) *RVs each taking values in the set $\{0, 1\}$ with the following joint probabilities:*

	$X_2 = 0$	$X_2 = 1$	
$X_1 = 0$	1/4	1/4	1/2
$X_1 = 1$	1/4	1/4	1/2
	1/2	1/2	1

From the above Table we can read for instance that the joint probability $\mathbf{P}((X_1, X_2) = (0, 0)) = 1/4$ and that the marginal probability $\mathbf{P}(X_1 = 0) = 1/2$.

By $x := y$ we mean that x is equal to y by definition. Let us consider the natural two-dimensional random vector (R$\vec{\text{V}}$) analogue in the real plane $\mathbb{R}^2 := (-\infty, \infty)^2 := (-\infty, \infty) \times (-\infty, \infty)$ of the Bernoulli(θ_1) random variable with probability $\theta_1 \in [0, 1]$ of realizing 1. A natural possibility is to use the **ortho-normal basis vectors** in $\mathbb{R}^2$:

$$e_1 := (1, 0), \qquad e_2 := (0, 1) \ .$$

This representation utilises the laws of vector addition and subtraction, which are performed component-wise, i.e. $(x_1, x_2) \pm (y_1, y_2) = (x_1 \pm y_1, x_2 \pm y_2)$. Let $\mathbb{1}_A(x)$ be the indicator function of set A, i.e. $\mathbb{1}_A(x) = 1$ if $x \in A$ and $\mathbb{1}_A(x) = 0$ if $x \notin A$.

Model 1 (Bernoulli(θ_1, θ_2) R$\vec{\text{V}}$). *Given a parameter $\theta_1 \in [0, 1]$ and $\theta_2 = 1 - \theta_1$, we say that $X := (X_1, X_2)$ is a* Bernoulli(θ_1, θ_2) *random vector (R$\vec{\text{V}}$) if it has only two possible outcomes from the set $\{e_1, e_2\} \subset \mathbb{R}^2$, i.e. $x := (x_1, x_2) \in \{(1, 0), (0, 1)\}$. The probability mass function (PMF) of the R$\vec{\text{V}}$ $X := (X_1, X_2)$ with realisation $x := (x_1, x_2)$ is:*

$$f(x; \theta_1, \theta_2) := \mathbf{P}(X = x) = \theta_1 \mathbb{1}_{\{e_1\}}(x) + \theta_2 \mathbb{1}_{\{e_2\}}(x) = \begin{cases} \theta_1 & if \quad x = e_1 := (1, 0) \\ \theta_2 = 1 - \theta_1 & if \quad x = e_2 := (0, 1) \\ 0 & otherwise \end{cases} \ .$$

We can write the Binomial(n, θ_1) RV Y_1 as a Binomial(n, θ_1, θ_2) R$\vec{\text{V}}$ $Y := (Y_1, Y_2)$, where $Y_2 = n - Y_1$. The parameters $\theta_1 + \theta_2 = 1$ and $\theta_1, \theta_2 \in [0, 1]$. In the coin-tossing context, this can be thought of keeping count of the number of `Heads` using Y_1 and the number of `Tails` using Y_2 in an IID sequence of n tosses of a coin with probability θ_1 of observing `Heads`. In general, by means of the ordered pair (Y_1, Y_2), we keep count of the two possible outcomes in an IID sequence of n trials. In the context of Galton's Quincunx this amounts to keeping count of the number of right and left turns made by each ball as it drops through n levels of pegs where the probability of a right turn at each peg is independently and identically θ_1.

Figure 2: Quincunx on the Cartesian plane. *The outline of the columns of shot that accumulate in the successive compartments approximates to the Curve of Frequency* (Galton 1889).

(a) Ten samples   (b) Thousand samples
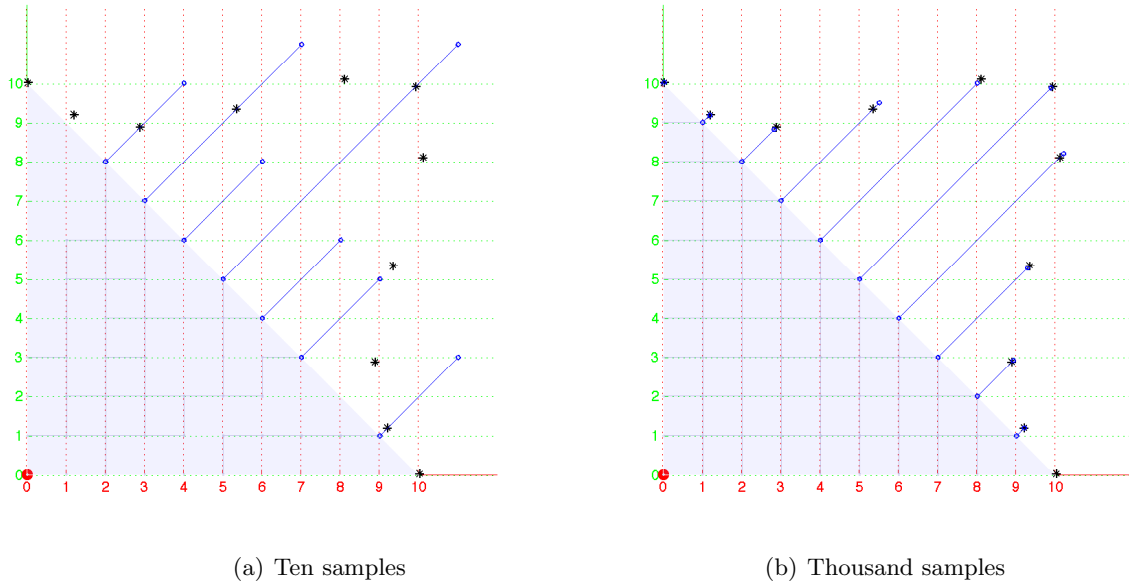
Model 2 (Binomial(n, θ_1, θ_2) R$\vec{\text{V}}$)**.** *The* Binomial(n, θ_1, θ_2) R$\vec{\text{V}}$ $Y := (Y_1, Y_2)$ *is the sum of* n *IID* Bernoulli(θ_1, θ_2) R$\vec{\text{V}}$s $X_1 := (X_{1,1}, X_{1,2}), X_2 := (X_{2,1}, X_{2,2}), \ldots, X_n := (X_{n,1}, X_{n,2})$, *as follows:*

$$Y := (Y_1, Y_2) = X_1 + X_2 + \cdots + X_n = (X_{1,1}, X_{1,2}) + (X_{2,1}, X_{2,2}) + \cdots + (X_{n,1}, X_{n,2})$$

$$= \sum_{i=1}^{n} (X_{i,1}, X_{i,2}) \ .$$

Binomial(n, θ_1, θ_2) R$\vec{\text{V}}$ $Y := (Y_1, Y_2)$, *with* $Y_2 = n - Y_1$ *and* $\theta_2 = 1 - \theta_1$, *has the PMF given by:*

$$f((y_1, y_2); n, \theta_1, \theta_2) := \mathbf{P}((Y_1, Y_2) = (y_1, y_2); n, \theta_1, \theta_2) = \binom{n}{y_1} \theta_1^{y_1} (1 - \theta_1)^{n - y_1} \ ,$$

where, the binomial coefficient that gives the number of distinct paths from the origin $(0, 0)$ *to the sample point* $(y_1, y_2) = (y_1, n - y_1)$ *in the Quincunx of Figure 2 (with* $n = 10$) *is:*

$$\binom{n}{y_1} := \frac{n!}{y_1!(n - y_1)!} \ .$$

In our depiction of Galton's Quincunx in Cartesian coordinates, as done in Figure 2, the top of the original mechanical Quincunx, the release point of each ball or shot, is mapped by a 90 degree rotation to the origin $(0, 0)$ to conform to the standard ortho-normal two-dimensional

coordinate system. Figure 2 may be interpreted as simulations of a Binomial($n = 10, \theta_1 = 0.5, \theta_2 = 0.5$) R$\vec{\text{V}}$ as the ordered pair resulting from the culmination of sample trajectories formed by the accumulating sum of $n = 10$ IID Bernoulli($\theta_1 = 0.5, \theta_2 = 0.5$) random vectors with outcomes, $(1,0)$, and $(0,1)$, according to probabilities, θ_1, and $\theta_2 = 1 - \theta_1$, respectively. The blue stem plots and black asterisks perpendicular to and above the diagonal line, i.e. the line connecting $(0,10)$ and $(10,0)$, are the density histogram of the samples and the PMF of our Binomial($n = 10, \theta_1 = 0.5, \theta_2 = 0.5$) R$\vec{\text{V}}$, respectively.

Note that with only 10 sample trajectories of 10 independent balls in Figure 2(a), the density histogram of the samples depicted by blue stem plots are different from the PMF of the Binomial($n = 10, \theta_1 = 0.5, \theta_2 = 0.5$) R$\vec{\text{V}}$ depicted by black asterisks. However, in Figure 2(b), with $1,000$ sample trajectories, we can see that the stem plots are reaching the asterisks to illustrate Galton's observation: *The outline of the columns of shot that accumulate in the successive compartments approximates to the Curve of Frequency. The Curve of Frequency* referred to by Galton here is the PMF of the Binomial($n = 10, \theta_1 = 0.5, \theta_2 = 0.5$) R$\vec{\text{V}}$ denoted by black asterisks and his reference to *The outline of the columns of shot that accumulate in the successive compartments* is merely the outline formed by the top of our blue stem plots of the density histograms. This phenomenon is formally known as convergence in distribution.

We can extend the Binomial(n, θ_1) RV Y_1 or the Binomial(n, θ_1, θ_2) R$\vec{\text{V}}$ $Y := (Y_1, Y_2)$ to its multivariate extension, namely, the Multinomial($n, \theta_1, \theta_2, \ldots, \theta_k$) R$\vec{\text{V}}$ $Y := (Y_1, Y_2, \ldots, Y_k)$. The multinomial distribution can be thought of as the outcome from n independent and identical trials where each trial results in exactly one of some fixed finite number k of possible outcomes and the probability $f(i)$ of outcome $i \in \{1, 2, \ldots, k\}$ in any one trial is θ_i, such that, $\sum_{i=1}^{k} \theta_i = 1$ and $f(i) = 0$ for each $i \notin \{1, 2, \ldots, k\}$. In other words, the multinomial random vector parameterised by (n, θ), where, $\theta := (\theta_1, \theta_2, \ldots, \theta_k)$, is $Y := (Y_1, Y_2, \ldots, Y_k)$, such that, each Y_i is the number of times that outcome $i \in \{1, 2, \ldots, k\}$ is observed in the n independent and identical trials. Thus, the Binomial(n, θ_1, θ_2) R$\vec{\text{V}}$ $Y := (Y_1, Y_2)$ is nothing but the Multinomial(n, θ_1, θ_2) R$\vec{\text{V}}$ $Y = (Y_1, Y_2)$, i.e. the binomial is the two-dimensional case ($k = 2$) of the multinomial.

This description of the multinomial distribution as the general case of the binomial is, of course, nothing new. What is less often seen is the equivalent of Model 2 (which showed how the binomial distribution can be developed as the sum of n IID Bernoulli(θ_1, θ_2) R$\vec{\text{V}}$s) for the multinomial distribution. In Models 3 and 4 below we develop the multinomial random vector as the sum of n IID de Moivre($\theta_1, \theta_2, \ldots, \theta_k$) R$\vec{\text{V}}$s, the natural multi-dimensional extension of the Bernoulli(θ_1, θ_2) R$\vec{\text{V}}$s.

Model 3 (de Moivre($\theta_1, \theta_2, \ldots, \theta_k$) R$\vec{\text{V}}$). *The PMF of the* de Moivre($\theta_1, \theta_2, \ldots, \theta_k$) *random vector* $X := (X_1, X_2, \ldots, X_k)$ *taking value* $x := (x_1, x_2, \ldots, x_k) \in \{e_1, e_2, \ldots, e_k\}$, *where the* e_i*'s are ortho-normal basis vectors in* $\mathbb{R}^k$ *is:*

$$f(x; \theta_1, \theta_2, \ldots, \theta_k) := \mathbf{P}(X = x) = \sum_{i=1}^{k} \theta_i \mathbb{1}_{\{e_i\}}(x) := \begin{cases} \theta_1 & if \quad x = e_1 := (1, 0, \ldots, 0) \in \mathbb{R}^k \\ \theta_2 & if \quad x = e_2 := (0, 1, \ldots, 0) \in \mathbb{R}^k \\ \vdots & \\ \theta_k & if \quad x = e_k := (0, 0, \ldots, 1) \in \mathbb{R}^k \\ 0 & otherwise \end{cases},$$

where, $\sum_{i=1}^{k} \theta_i = 1$.

Addition of n IID de Moivre$(\theta_1, \theta_2, \ldots, \theta_k)$ R$\vec{V}$s yields the Multinomial$(n, \theta_1, \theta_2, \ldots, \theta_k)$ random vector that is defined below.

Model 4 (Multinomial$(n, \theta_1, \theta_2, \ldots, \theta_k)$ R$\vec{V}$). *The* Multinomial$(n, \theta_1, \theta_2, \ldots, \theta_k)$ R$\vec{V}$ $Y :=$ $(Y_1, Y_2, \ldots, Y_k)$ *is the sum of n IID* de Moivre$(\theta_1, \theta_2, \ldots, \theta_k)$ R$\vec{V}$s:

$$X_1 := (X_{1,1}, X_{1,2}, \ldots, X_{1,k}), X_2 := (X_{2,1}, X_{2,2}, \ldots, X_{2,k}), \ldots, X_n := (X_{n,1}, X_{n,2}, \ldots, X_{n,k}) \ ,$$

as follows:

$$\begin{aligned}
Y &:= (Y_1, Y_2, \ldots, Y_k) \\
&= X_1 + X_2 + \cdots + X_n \\
&= (X_{1,1}, X_{1,2}, \ldots, X_{1,k}) + (X_{2,1}, X_{2,2}, \ldots, X_{2,k}) + \cdots + (X_{n,1}, X_{n,2}, \ldots, X_{n,k}) \\
&= \sum_{i=1}^{n} (X_{i,1}, X_{i,2}, \ldots, X_{i,k}) \ .
\end{aligned}$$

Note that any realization $y := (y_1, y_2, \ldots, y_k)$ of our R$\vec{V}$ $Y := (Y_1, Y_2, \ldots, Y_k)$ is in the set $\mathbb{Y}$, a $(k-1)$-simplex:

$$y := (y_1, y_2, \ldots, y_k) \in \mathbb{Y} := \left\{ (y_1, y_2, \ldots, y_k) : y_1, y_2, \ldots, y_k \in \{0, 1, 2, \ldots\}, \sum_{i=1}^{k} y_i = n \right\} \ ,$$

Thus $y := (y_1, y_2, \ldots, y_k)$ is a k-dimensional non-negative integer vector whose components sum to n and the PMF of Y is given by:

$$f(y; n, \theta) := f(y; n, \theta_1, \theta_2, \ldots, \theta_k) := \mathbf{P}(Y = y; n, \theta_1, \theta_2, \ldots, \theta_k) = \binom{n}{y_1, y_2, \ldots, y_k} \prod_{i=1}^{k} \theta_i^{y_i} \ ,$$

where the multinomial coefficient that gives the number of distinct paths from the origin $(0, 0, \ldots, 0)$ to the sample point $(y_1, y_2, \ldots, y_k) \in \mathbb{Y}$ is:

$$\binom{n}{y_1, y_2, \ldots, y_k} := \frac{n!}{y_1! y_2! \cdots y_k!} \ .$$

We can visualise the construction of the Multinomial$(n, \theta_1, \theta_2, \theta_3)$ random vector as a sum of n IID de Moivre$(\theta_1, \theta_2, \theta_3)$ R$\vec{V}$s via a three dimensional extension of the Quincunx called the "Septcunx" and relate the number of paths that lead to a given trivariate sum (y_1, y_2, y_3) with $\sum_{i=1}^{3} y_i = n$ as the multinomial coefficient $\frac{n!}{y_1! y_2! y_3!}$. In the Septcunx, balls choose from one of three paths along e_1, e_2 and e_3 with probabilities θ_1, θ_2 and θ_3, respectively, in an IID manner at each of the n levels, before they collect at buckets placed at the integral points in the 2-simplex, $\mathbb{Y} = \{(y_1, y_2, y_3) \in \mathbb{Z}_+^3 : \sum_{i=1}^{3} y_i = n\}$. Once again, we can visualise that the sum of n IID de Moivre$(\theta_1, \theta_2, \theta_3)$ R$\vec{V}$s constitute the Multinomial$(n, \theta_1, \theta_2, \theta_3)$ R$\vec{V}$ as depicted in Figure 3. For simplicity, in this project the multinomials used have $\theta_i = \frac{1}{k}$, i.e. each of the $i \in \{1, 2, \ldots, k\}$ outcomes of a single trial is equally likely. When $k = 3$, the Multinomial$(n, \theta_1, \theta_2, \theta_3)$ R$\vec{V}$ $Y := (Y_1, Y_2, Y_3)$ is known a the trinomial R$\vec{V}$.

Figure 3: Septcunx on the Cartesian co-ordinates. One thousand simulated samples of the Multinomial($n, \theta_1 = 1/3, \theta_2 = 1/3, \theta_3 = 1/3$) R$\vec{V}$ as the sum of n IID de Moivre($\theta_1 = 1/3, \theta_2 = 1/3, \theta_3 = 1/3$) R$\vec{V}$s over $\{(1,0,0),(0,1,0),(0,0,1)\}$ with probabilities $\{\theta_1, \theta_2, \theta_3\}$, respectively are shown in subfigure (a) for $n = 2$ and in subfigure (b) for $n = 10$. The blue lines perpendicular to the sample space of the Multinomial($n, \theta_1, \theta_2, \theta_3$) R$\vec{V}$, i.e. the triangular plane in $\mathbb{R}^3$ connecting $(n,0,0)$, $(0,n,0)$ and $(0,0,n)$, are the density histogram of the one thousand samples.
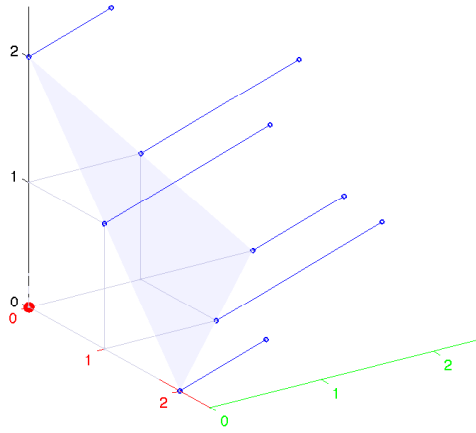
(a) Thousand Samples with $n = 2$
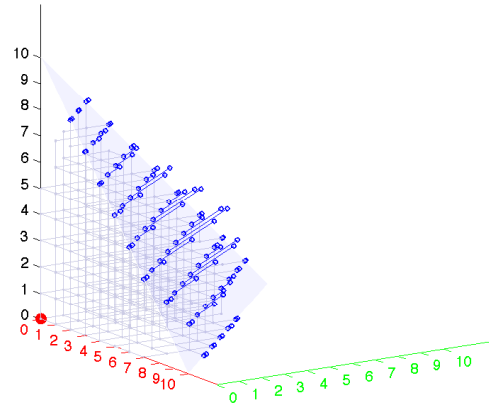
(b) Thousand Samples with $n = 10$

3. Visual Cognitive Tool for Quincunx and Septcunx

The VCT GUI requires a way to represent the binomial and trinomial distribution visually. In this project, a directional travel analogy is used. The sequence of trials is represented as a journey in which the outcome of any of the n trials is interpreted as the direction taken at that point (out of k possible directions) and the final outcome can be interpreted in terms of an end point of the journey in relation to the starting point. A multinomial random vector $Y := (Y_1, Y_2, \ldots, Y_k)$ is the summative effect of n distinct independent and identically distributed directional movements. The direction analogy is interpreted as travel in orthogonal directions. Thus, the multinomial random vector $Y := (Y_1, Y_2, \ldots, Y_k)$ is considered as the sum of n independent and identically distributed trials in which the result of any one trial is e_i, the i-th orthonormal basis vector (the k-dimensional vector where all elements except the i-th are 0 and the i-th element is 1).

In keeping with the Quincunx approach, the GUI depicts a sample, when in motion (i.e. accumulating its trials, or the equivalent of dropping from peg to peg for the Quincunx), as a ball. The GUI represents the 'path' taken by a particular sample or ball. For any trial

(or level or stage of the journey) m, such that $1 \leq m \leq n$, a sample $y = (y_1, y_2)$ can be thought of as being at the point represented by $\sum_{j=1}^{m} x_j$, where $x_j = (x_{j,1}, x_{j,2})$ is one of the ortho-normal basis vectors that is encoding the outcome of the j-th trial (i.e. x_j is either $(1,0)$ or $(0,1)$). This point can be plotted on the two-dimensional Cartesian "(x,y)" coordinates for each $m = 1, 2, \ldots, n$ and thus the 'path' of the sample plotted. This demonstrates that different paths can lead to the same outcome in the sample space and provides a visual interpretation of the binomial coefficient $\binom{n}{y_1}$ as the number of distinct paths that can lead to the same destination point at level n, namely $y = (y_1, y_2) = (y_1, n - y_1) = \sum_{j=1}^{n}(x_{j,1}, x_{j,2})$. The model used in the project assumes that $\theta = (\frac{1}{2}, \frac{1}{2})$, i.e. in any trial, travel in either one of the two orthogonal directions is equally likely.

The same approach can be taken when $k = 3$, the trinomial case, by plotting outcomes on the three-dimensional "(x, y, z)" Cartesian coordinates. This works reasonably well and provides a way to visualise the sample space of a Multinomial$(n, \theta_1, \theta_2, \theta_3)$ R$\vec{\text{V}}$ as $(n+1) \times (n+2) \div 2$ points on a 2-simplex (triangle) on a sloping plane. For the project GUI, $\theta = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, i.e. in any trial, travel in the "x", "y" or "z" directions is equally likely. The representations on Cartesian coordinates of the sample space for a multinomial random vector with number of trials $n = 10$ are shown below in Figure 4 for $k = 2$ and $k = 3$. The three-dimensional version of the Quincunx has been termed the Septcunx, along Galton's Latin nomenclature concerning the geometry of the graph structure underpinning the probability model.

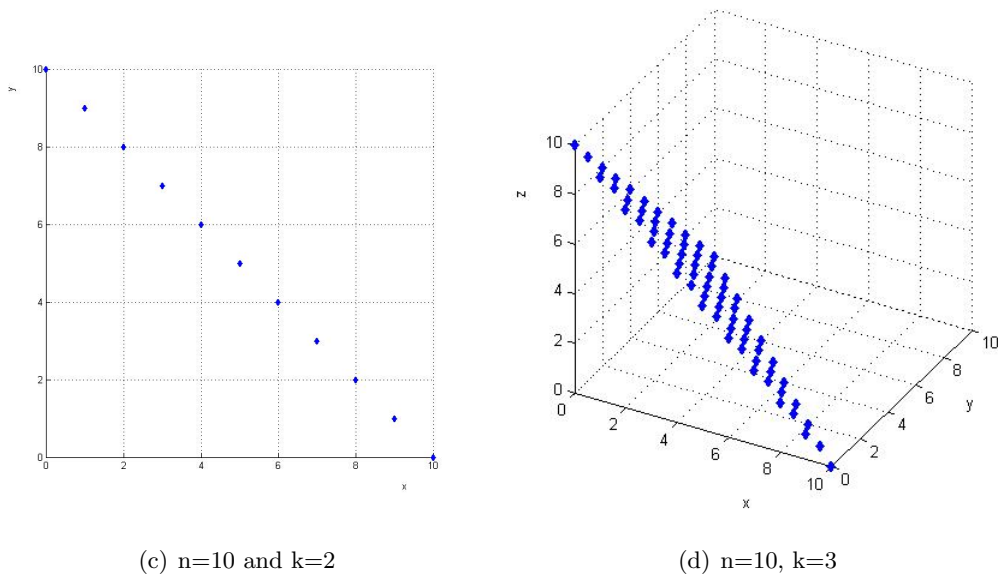(c) n=10 and k=2   (d) n=10, k=3

Figure 4: Binomial and trinomial sample spaces represented on Cartesian coordinates in (a) and (b), respectively.

The trinomial GUI should convey information about the probability mass function (PMF), i.e. we want to plot an outcome in the sample space in such a way that it also gives some sense of the probability of that outcome. This is achieved by scaled orthogonal projections from the plane containing the sample space. The length of the projection is proportional to the PMF for that outcome and to the number of trials n. Similarly, the proportion of realisations

having a particular outcome can be represented by a version of the stem plot, or actually drawing the projected line from the representation of the outcome on the sample space plane and giving it a length proportional to the proportion of realisations having that outcome. The representations on Cartesian coordinates of the sample space with PMF-scaling for a multinomial random vector with number of trials $n = 10$ are shown below in Figure 5 for $k = 2$ and $k = 3$.

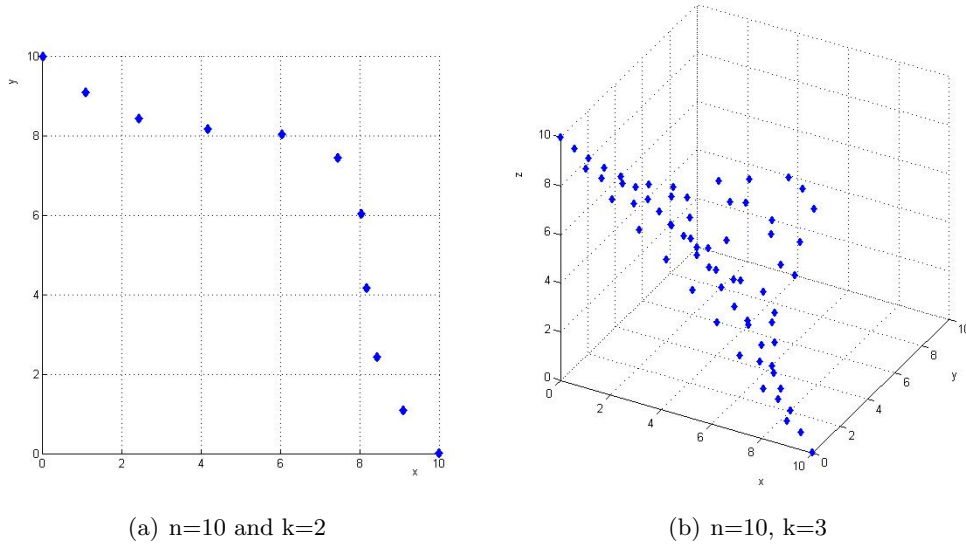(a) n=10 and k=2   (b) n=10, k=3

Figure 5: Binomial and trinomial PMF indicated by scaled orthogonal projection over the sample space in (a) and (b), respectively.

We use screen shots of our GUI in action to illustrate the results of the project. This GUI is available for download from http://www.math.canterbury.ac.nz/~r.sainudiin/codes/ quinseptcunx/. The Matlab source code for our GUI is included as associated files. These files are commented to allow extensions of the GUI that are mentioned in § 5. The source code is licensed under the Creative Commons Attribution-Noncommercial-Share Alike 3.0 New Zealand Licence. To view a copy of this license, visit http://creativecommons.org/ licenses/by-nc-sa/3.0/nz/.

Figure 6 shows the Multnomial GUI in Quincunx ($k = 2$) mode, simulating first one and then 20 realisations from a binomial random vector with $n = 10$, $\theta = (\frac{1}{2}, \frac{1}{2})$. The range of the sample space is indicated with a light blue simplex (which is a 1-simplex or line for two-dimensional coordinates (x, y)), and the area under the simplex is also filled in light blue. The outcomes in the sample space, scaled as described above, are represented with a pentagram-star and the proportion of the realisations resulting in each outcome in the sample space is indicated with the stem plot. The 'history path' or trajectory of this random walk from the origin $(0, 0)$ to its final destination or outcome in the sample space is shown as a light grey line below the simplex.

Analogous screenshots for the Multnomial GUI in Septcunx ($k = 3$) mode are shown below in Figure 7, simulating first one and then 50 realisations from the trinomial random vector with $n = 10$, $\theta = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. The sample space is now the set of three dimensional non-negative
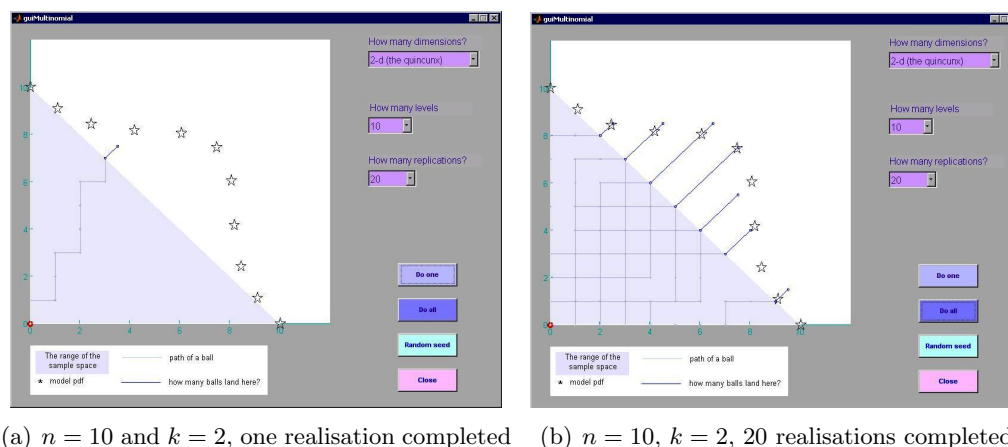
(a) $n = 10$ and $k = 2$, one realisation completed   (b) $n = 10$, $k = 2$, 20 realisations completed

Figure 6: The Multinomial GUI in Quincunx mode.

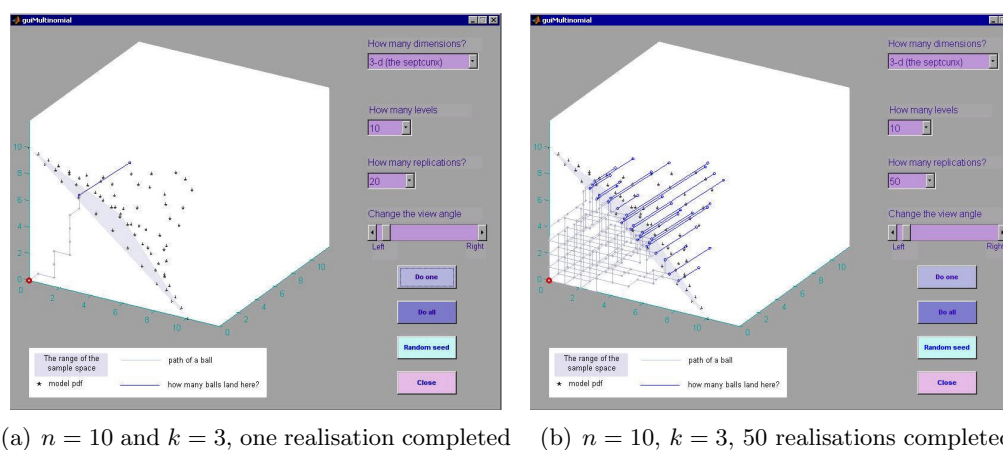integer vectors that lie on the 2-simplex or triangle whose points sum to n.

(a) $n = 10$ and $k = 3$, one realisation completed   (b) $n = 10$, $k = 3$, 50 realisations completed

Figure 7: The Multinomial GUI in Septcunx mode.

When the GUI is put into Septcunx mode a default viewing position is provided which angles the view to try to maximise the three-dimensional effect that the GUI aims for. The user is also provided with an additional control in the form of a slider bar which allows her/him to interactively rotate the Septcunx plot through approximately 60 degrees about the vertical axis. Figure 8 below shows the effect of rotating the view of the Septcunx using the slider.

4. Kinesthetic and Experimental Learning with Physical Models

A physical model of the Quincunx that was constructed in Ashman and Lawrence (2007) is introduced to the students before the VCT Quincunx GUI. Students are invited to drop a ball into the top of the Quincunx with their hands and asked to observe the effect their drop (however carefully controlled) has on the trajectory and the final outcome of the ball. They are asked to drop more than one ball to make comparisons between the realisations.

(a) Part-way through rotation   (b) Towards the end of a rotation
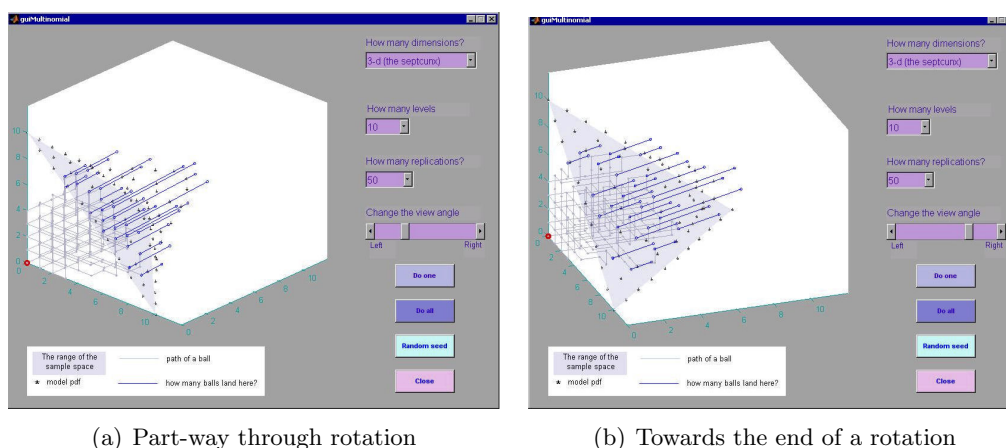
Figure 8: The Multinomial GUI in Septcunx mode with $n = 10$, $k = 3$, and 50 replications completed.

Such direct observations of balls that are released from a student's own hands, as they are dropping through the Quincunx, are no substitute for the VCT GUI. Since this activity can take some time, the Quincunx is left in a freely accessible area where the students can experiment, play and learn in their own time. The sensitivity in initial conditions during the neuro-muscular controlled hand-release of the ball is pointed out as the source of randomness that yields different trajectories. This is appreciated by students who have had some exposure to dynamical systems and nonlinear differential equations.

However, formal mathematical constructions of binomial and trinomial random vectors that are communicated in the traditional read/write style of teaching/learning can be visually reinforced using our VCT GUIs. This interactive visual reinforcement of the mathematical probability models by means of our visual cognitive tool GUI is the primary emphasis of this paper.

In order to kinesthetically experience and learn a physical Septcunx lattice, students are optionally provided with a chemistry molecule-building kit with tetrahedral-connectors and connecting-tubes that are labelled by $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$. A physical model of the Septcunx with balls dropping through the hollow tubes of the same length that are connected by hollow tetrahedral junctions is possible if the angles made by the three tubes below each junction are symmetrically obtuse as in the crystallographic model for the arrangement of carbon atoms in diamond. This allows the balls to drop down to the next level due to gravity. Students are invited to assemble such a physical Septcunx lattice with their own hands from the connectors and tubes. This activity is meant to kinesthetically reinforce the visually interactive animations and graphics encountered with the VCT Septcunx GUI.

5. Conclusion and Potential Modification

This project has explored how binomial and trinomial GUIs based on vector addition can be used as visual cognitive tools (VCTs) to enhance student learning. The binomial and trinomial
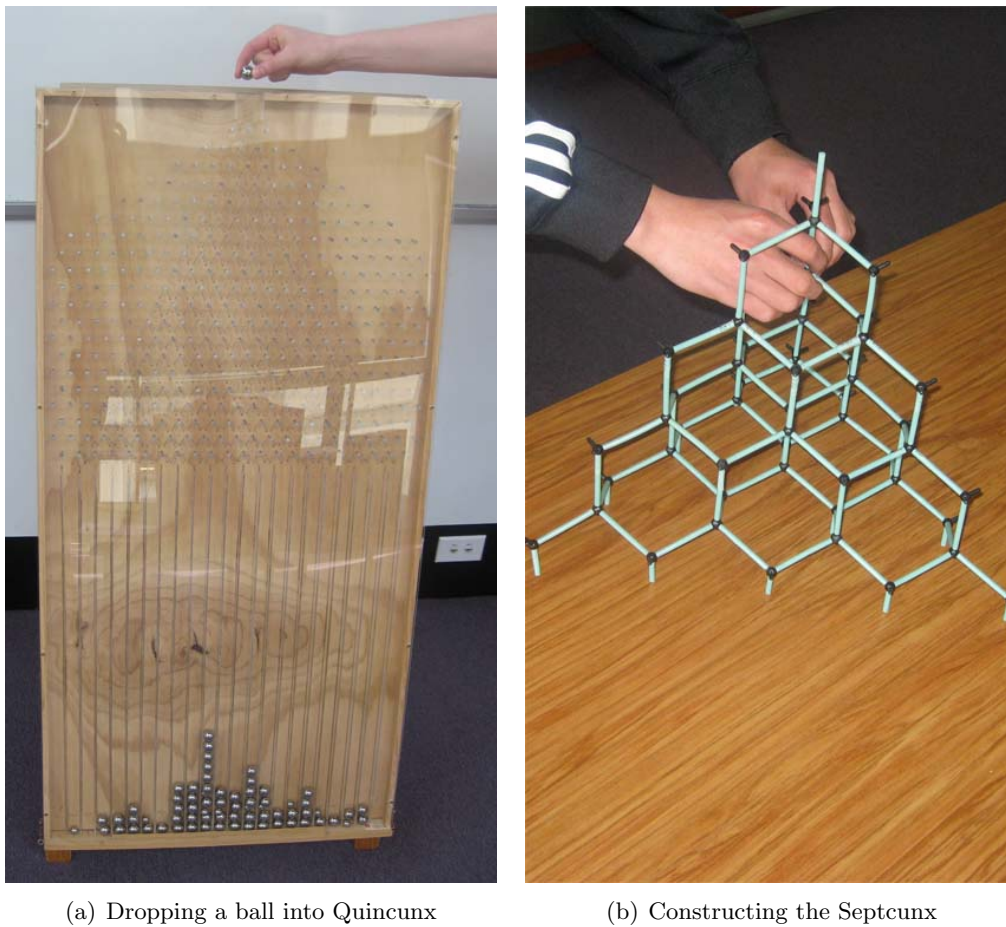
(a) Dropping a ball into Quincunx (b) Constructing the Septcunx

Figure 9: A student is about to release her second ball into our physical Quincunx and another is constructing the Septcunx lattice.

distributions are visually presented from their Quincunx and Septcunx constructions along with their corresponding combinatorial coefficients as the number of distinct paths to reach a given sample point. The project GUIs do this by adapting graphical displays in the following ways:

- Break processes into steps with illustrations of each stage, highlighting important points;

- Allow the user to compare the effects of using different parameters;

- Allow the user controlled interaction with the data and displays;

- Provide the graphical data displays for the user so that the user is not distracted by the mechanics of creation.

The binomial and trinomial GUI differs from the traditional visualisations which seem to be available to students in at least five ways:

1. The GUI emphasises the underlying random vector model for the Bernoulli, binomial, de Moivre and multinomial. This treatment is more consistent and general than the more common presentation of the Bernoulli as a "trial" which may result in success or failure, and then a binomial random variable as the sum of the successes in n Bernoulli trials, and finally the introduction of a vector of outcomes when the multinomial is described as a generalisation of the binomial to a situation where a trial may have more than two outcomes.

2. The GUI allows the student to see how the binomial and trinomial arise as the sums of Bernoulli and three-dimensional de Moivre random vectors, respectively. Many other visual presentations of binomial concentrate on showing how the binomial reacts to changes in n and p. Visualisations of the relationship of the binomial to the Bernoulli also exist, including visualisations using the Galton Quincunx, but do not extend this to the multinomial. In fact, visual treatments of the multinomial for students are relatively uncommon and examples which attempt to model the multinomial as a sum of other random variables/vectors even more so.

3. The GUI, like Galton's Quincunx, enables the student to understand the curve of frequency and hence the binomial coefficient by showing the binomial arising as the sum of n Bernoulli random vectors. In particular, the GUI shows not just the 'path' of the latest or current ball but also the 'history paths' of previous balls. Other visualisations of the Galton Quincunx seem only to show the latest ball. The history paths are important in helping the student to understand how many different paths can lead to each of the possible end results. When the trinomial version is shown, the student can then see how naturally the two dimensional result extends to three dimensions, how the 'bell' (rather than curve) of frequency arises, and hence gain insight into the general multinomial coefficient.

4. The GUI makes maximum use of human capabilities for visualising dimensionality. It does this by mapping random vector dimensions to Cartesian coordinates ("(x, y)" for the binomial, "(x, y, z)" for the trinomial) and then using projections from a sample space represented as a simplex on a sloping plane (a 1-simplex or line for the binomial, a 2-simplex or triangle for the trinomial) to represent the relative frequency of actual realisations as well as the probability mass function. Other representations usually take a bar-chart or histogram approach, using height to represent frequency or probability. This approach would struggle to show both the three-dimensional nature of a trinomial random vector and at the same time the curve of frequency, so that even if the binomial is shown as two-dimensional random vector with a curve of frequency in the third dimension, the trinomial cannot be presented consistently. The GUI discussed in this paper uses a consistent presentation of the binomial and trinomial, which again allows the student to see more clearly how these distributions are related as they arise from random vector addition.

5. In our Binomial and trinomial models we set $(\theta_1, \theta_2) = (1/2, 1/2)$ and $(\theta_1, \theta_2, \theta_3) = (1/3, 1/3, 1/3)$, respectively, in order to focus on the visual understanding of the simpler equi-probable models. Once this understanding has been achieved it is possible to change the source code of our VCT to set (θ_1, θ_2) and $(\theta_1, \theta_2, \theta_3)$ to be arbitrary probabilities.

GUI design and human interaction technologies are specialised fields in their own right. This project is intended as a demonstration of the potential for GUIs as VCTs, and the actual GUIs created probably violate most if not all of the rules for good GUI design. Considerable improvements could be made on the design of these GUIs as human/technology interfaces. In addition, the programming could undoubtedly be improved and made cleaner and more efficient.

The major disadvantage of the Cartesian coordinate representation is that it is not good for representing $k > 3$ Multinomial $\vec{\text{RV}}$s. An alternative representation which retains some aspects of the travel/direction analogy but which can be adapted for higher dimensions is the use of orbit plots or tours. An orbit plot uses multiple axes and projects a high-dimensional representation into a number of low-order representations on the different axes (Young, Valero-Mora and Friendly, 2006). This would be interesting to explore but is beyond the scope of this project.

More ways in which the user can move and rotate the plot of trinomial outcomes could be added to the Septcunx GUI. More emphasis could be given to the meaning of the multinomial coefficient $\binom{n}{y_1, y_2, \ldots y_k} = \frac{n!}{y_1! y_2! \ldots y_k!}$, possibly by tracking how many different paths in the Cartesian representation of the multinomial can lead to each outcome.

The GUI could also be expanded, or different versions created, to illustrate other related aspects. For example, with our GUI as a visual cognitive tool, one could emulate Galton's original purpose of the Quincunx as well as the Septcunx, its three-dimensional extension. These include the Bernoulli's Law of Large Numbers and the Central Limit Theorem. The visually interactive animation of the Law of Large Numbers (LLN) and the Central Limit Theorem (CLT) are possible in a mathematically rigourous manner for two and three dimensions with our VCT since our constructive definitions of the binomial and trinomial random vectors have appropriately embedded the probability models in the real Euclidean spaces of dimension two and three. Thus we can animate fundamental definitions such as convergence of random variables and random vectors that are needed to appreciate LLN and CLT for our probability models, example cases of two basic asymptotic theorems in statistics.

We have also used the physical model of our Quincunx to illustrate basic likelihood concepts, including maximum likelihood estimation of the probability of the right turn as in Ashman and Lawrence (2007). With some additional work, one can provide a visual derivation of the limiting Poisson approximation of the biased binomial distribution as well as an animated interactive derivation of Pearson's Chi-sqared distribution with 1 or 2 degrees of freedom as done in Pearson (1900) along the arguments given in Knuth (1998) by starting from our Model 4 with $k = 2$ or 3. Thus our VCT makes a significant step in the general direction of rigourously animating special cases of fundamental statistical theorems in a visually interactive fashion for students of the "YouTube" generation as encouraged by Professor David Spiegelhalter of Statistical Laboratory, Cambridge University, on the 17th of July 2008 during the 7th World Congress in Probability and Statistics that was held in Singapore.

Acknowledgements

6. References

Ashman, B. and Lawrence, R. (2007). Estimating the binomial probability p for Galton's Quincunx. In *Computational statistical experiments: STAT218-07S2(C) Student Projects Report UCDMS 2008/5, 47–51*. Retrieved Oct. 4, 2008, from http://www.math.canterbury.ac.nz/~r.sainudiin/courses/STAT218/projects/Stat218StudentProjects2007.pdf.

Card, S.K., Mackinley, J.D. and Shneiderman, B. Eds. (1999). *Readings in information visualization: Using vision to think*. San Francisco, CA: Morgan Kaufmann.

Galton F. (1889). *Natural Inheritance*, Macmillan. Publicly available from http://www.galton.org/books/natural-inheritance/index.html

Knuth, D.E. (1998). *The Art of Computer Programming, Vol. 2 / Seminumerical Algorithms, 3rd Ed*, 55–56. London: Addison-Wesley.

Nolan, D. and Temple Lang, D. (2003). *Case Studies and computing: Broadening the scope of statistical education*. Retrieved October 14, 2008, from http://www.stat.berkley.edu/~nolan/Papers/isi03.pdf.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophical Magazine, 5th Series 50 157–175. Reprinted in *Karl PearsonÕs Early Statistical Papers* (1948) 339–357 Cambridge Univ. Press.

Sedig, K. and Liang, H. (2008). Learner-information interaction: A macro-level framework characterizing visual cognitive tools. *Journal of Interactive Learning Research*, 19(1) 147–173.

Tufte, E.R. (1983). *The visual display of quantitative information*. Cheshire, CN: Graphics Press.

Tukey, J.W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Young, F., W., Valero-More, P.M. and Friendly, M. (2006). *Visual Statistics: Seeing data with dynamic interactive graphics*. Hoboken, NJ: John Wiley & Sons.