Beyond the Quintessential Quincunx

by

Michael A. Proschan
National Institute of Allergy and Infectious Diseases

and

Jeffrey S. Rosenthal
Department of Statistics
University of Toronto

Technical Report No. 0905 September 8, 2009

TECHNICAL REPORT SERIES

University of Toronto

Department of Statistics

Beyond the Quintessential Quincunx

Michael A. Proschan [1] and Jeffrey S. Rosenthal [2]

ABSTRACT

The quincunx, a contraption with balls rolling through a triangle-shaped arrangement of nails, was invented to illustrate the binomial distribution and the central limit theorem for Bernoulli random variables. As it turns out, the quincunx can be used to teach many different concepts, including the central limit theorem for independent but not identically distributed random variables, permutation tests in a paired setting, and the generation of a random variable with an arbitrary continuous distribution from a uniform variate. This paper uses quincunx applets to illustrate these and other applications.

KEY WORDS: Bernoulli Random Variables; Binomial Distribution; Central Limit Theorem; Permutation Test; Random Number Generation; Symmetric Distribution

1 Introduction

The Quincunx is a device invented by Francis Galton to illustrate the binomial distribution and the central limit theorem (CLT) for Bernoulli random variables. For the very interesting history, see pages 275-281 of Stigler (1986) or Kunert, Montag, and Pohlmann (2001). In this device, balls roll one-by-one down a board, striking nails head-on and deflecting to the left or right, and then collect in bins at the bottom (Figure 1).

Each deflection is a binary random variable parameterized with left and right corresponding to -1 and $+1$, respectively. Regardless of the deflections in previous rows, the ball is equally likely to go left or right in the current and future rows, so the deflections in different rows are independent. The position of the ball at the end is determined by the sum of the deflections; with $2n$ rows, if there are just as many left as right deflections, the sum of the -1s and $+1$s is $S_{2n} = 0$ and the ball will end in the middle chute. On the other hand, if all of the deflections are to the right ($+1$), the sum will be $S_{2n} = 2n$ and the ball will end in the rightmost chute, etc. Each ball is a single realization of S_{2n}, whereas the proportions of balls in the different chutes represent the simulated distribution of S_{2n}. With a large number of balls, the simulated distribution will tend to the true distribution of S_{2n}. If we relabel left and right deflections as 0 and 1 instead of -1 and $+1$, the distribution of S_{2n} is binomial $(2n, 1/2)$. With a large number of rows, this binomial distribution will look bell-shaped,

[1] National Institute of Allergy and Infectious Diseases (E-mail: *ProschaM@mail.nih.gov*)
[2] University of Toronto (E-mail: *jeff@math.toronto.edu*)

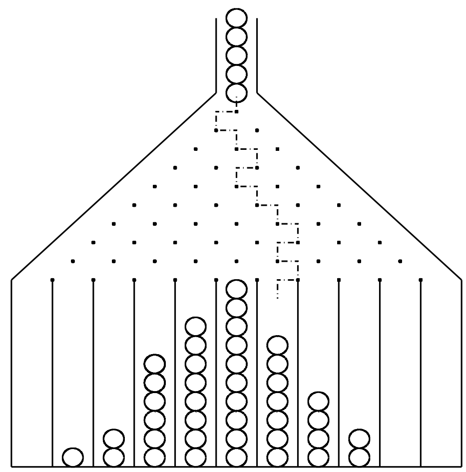

Figure 1: A quincunx. Balls roll down the board, striking one nail in each row and bouncing to the left or right with equal probability. The bell-shaped distribution of balls in the bins at the bottom simulates the binomial distribution and illustrates the CLT.

consistent with the CLT. Of course the quincunx shows only that *some* bell-shaped distribution applies, not necessarily the normal. For a heuristic justification of why only a normal distribution fits, see Proschan (2007).

Commercial quincunxes are scarce and expensive, and building even a simple one (Figure 2) is a painstaking endeavor because the nails must be placed in just the right places, the board must be very straight, and the balls must be the same size and not have imperfections. Building a universal quincunx that incorporates the modifications we make in this paper is nearly impossible. Fortunately, a universal quincunx –"uncunx"– applet is available at

$$\text{http://probability.ca/jeff/java/uncunx.html} \tag{1}$$

It shows triangles instead of nails to indicate where the balls branch to the left or right, and shows them only when the balls are actually branching. The reason for this feature is that the uncunx allows different deflection sizes in different rows, which markedly increases the required number of nails, cluttering the display if they were all shown simultaneously. Clicking on the display and typing "f" or "s" causes the balls to go faster or slower, respectively. As each ball reaches the bottom of the uncunx, the corresponding histogram bar becomes taller. Also shown is a normal curve with the same mean and variance as the binomial. As more balls accumulate, the histogram looks very similar to the normal curve. Later we will illustrate more sophisticated features of the uncunx.

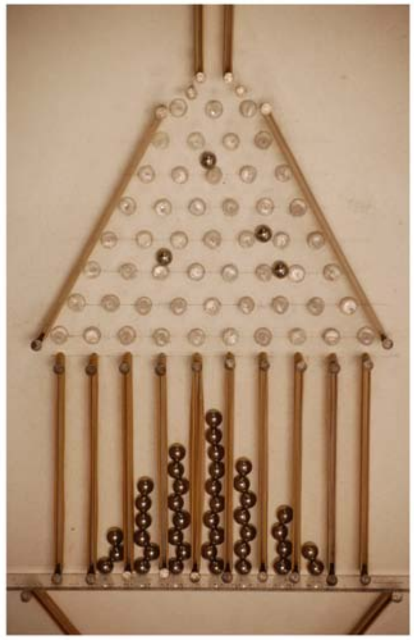

Figure 2: A homemade quincunx made with pushpins, rubber bands, a ruler, steel balls, nails, and a board.

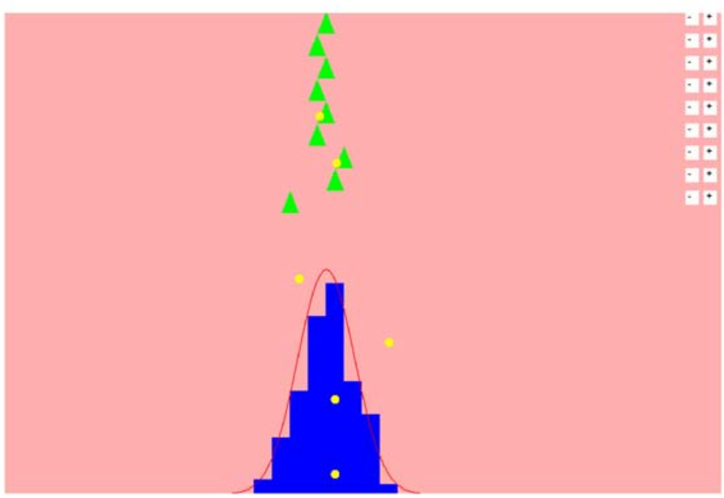

Figure 3: The uncunx applet in its default mode with equal row deflections. The superimposed normal distribution has the same mean and variance as the binomial it simulates.

Table 1: *Pascal's Triangle. Each element of a row other than the ones at the end is obtained by adding the two adjacent elements of the preceding row.*

				1					row 0
			1		1				row 1
		1		2		1			row 2
	1		3		3		1		row 3
1		4		6		4		1	row 4
⋮	⋮		⋮		⋮		⋮		⋮

This paper argues that the uncunx is a wonderful teaching device that can be used to illustrate many different concepts in probability and statistics, including Pascal's triangle, CLTs for independent but not identically distributed random variables, and permutation tests. The presentation is heuristic rather than rigorous, but the intuition instilled by the uncunx is valuable for statistics students of all levels.

2 Pascal's Triangle

One reason a quincunx is helpful is that it gets students to think about the number of paths a ball can take, which can help with combinatorics. Number the rows sequentially beginning with 0, and number the nails in each row sequentially beginning with nail 0. How many paths lead to nail k of row r of the quincunx? Denote this number by $N_{r,k}$. The number of paths is simple if $k = 0$; every deflection must be to the left, so $N_{r,0} = 1$. Similarly, $N_{r,r} = 1$. Now consider nail k of row r, where $k \neq 0$ and $k \neq r$. For the ball to strike nail k in row r, in row $r - 1$ it must have struck either nail $k - 1$ and bounced right, or nail k and bounced left. Therefore, the number of paths leading to nail k in row r is the sum of the numbers of paths striking nails $k - 1$ and k of row $r - 1$. In symbols, $N_{r,k} = N_{r-1,k-1} + N_{r-1,k}$ (see equation 1.11 on page 37 of Parzen, 1960). This suggests an iterative way to determine $N_{r,k}$. Place a 1 in row 0 and at the extreme left and right ends of each row. To get the number corresponding to a given nail in a row, add the numbers in the preceding row just to the left and just to the right of the position in the current row. This is the algorithm for Pascal's triangle (Table 1). The equation $N_{r,k} = N_{r-1,k-1} + N_{r-1,k}$ is usually proven by induction, which offers no intuition. The quincunx is a concrete model that provides the intuition.

3 Central Limit Theorems for Symmetric Binary Random Variables

The Introduction explained that the bell-shaped distribution of balls in the different chutes illustrates the CLT for iid symmetric Bernoulli random variables, but does the CLT hold for independent but not identically distributed symmetric binary random variables? Suppose that deflections in different rows have different fixed lengths, $|d_1|, \ldots, |d_n|$. Then $D_1, \ldots, D_n$ are independent binary random variable with $\Pr(D_i = +|d_i|) = \Pr(D_i = -|d_i|) = 1/2$. Will the sum $\sum_{i=1}^{n} D_i$ still be approximately normal? The applet at the URL (1) can be used to answer this question. The $+$ or $-$ boxes in the upper right portion of the applet enlarge or shrink the deflections in the different rows. The variance of the superimposed normal distribution, $\sum d_i^2$, changes as we change the $|d_i|$. Click on the $+$ box of the first row 15 times. The first deflection dominates all other rows, and the resulting bimodal histogram is clearly not approximately normal (Figure 4). The student can play with the deflection sizes to gain insight about how dissimilar they have to be for the resulting histogram to appear non-normal. From a practical standpoint, the applet simulation is more directly relevant than theoretical asymptotic results. After all, even if we can prove that S_n is asymptotically normal under certain conditions, n may need to be very large for the normal approximation to be accurate.

The bimodal distribution from the uncunx applet suggests a way to construct an infinite set of symmetric binary random variables such that S_n is not asymptotically normal. We can make the deflection in the first row dwarf the sum of all subsequent deflections. To make it easy to compute the sum of all subsequent absolute deflections, choose $|d_2|, |d_3|, |d_4| \ldots$ to be a geometric series, say $|d_2| = 1/4$, $|d_3| = 1/8$, $|d_4| = 1/16$ etc. Then $\sum_{i=2}^{\infty} |d_i| = (1/4)/(1-1/2) = 1/2$. Simply choose $|d_1|$ larger than $1/2$, say $|d_1| = 1$. This will create two separate humps in the distribution of the sum similar to what happened in Figure 4. For example, if $D_1 = -1$, then even if all of the remaining D_i are positive, $S_n \leq -1+1/2 = -1/2$. Similarly, if $D_1 = +1$, then even if all of the remaining D_i are negative, $S_n \geq 1 - 1/2 = 1/2$. Therefore, $P(-1/2 < S_n < 1/2) = 0$ for all n, and the distribution of S_n has two symmetric humps. The conditions ensuring asymptotic normality must therefore rule out cases like this in which one row's deflection dominates the overall sum. Students in undergraduate probability and statistics courses need not understand the precise Lindeberg condition that ensures normality (e.g., see page 239 of Feller, 1957), just that some condition is needed to

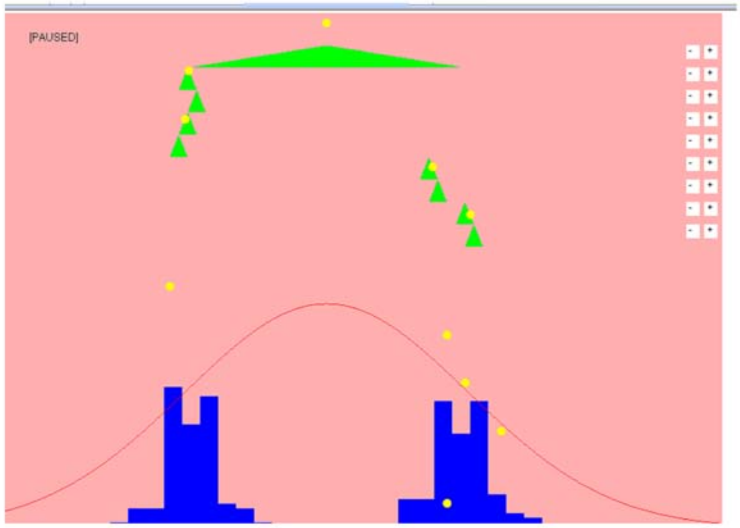

Figure 4: The applet after clicking on the "+" button in the first row 15 times to make its deflection dominate all others, causing a bimodal distribution.

prevent one row from dominating.

4 Permutation Tests in A Paired Setting

Sometimes data consist of pairs, one member of which is given a treatment (T) and the other a control (C). Examples are crossover studies or clinical trials with pair matching. Under the null hypothesis of no treatment effect, the distribution of paired differences should be symmetric about 0. One attractive option for analyzing such data is a permutation test (Good, 2005). We condition on the absolute value of pairwise differences, $|d_i|$, $i = 1, \ldots, n$ Conditioned on $|d_i|$, the ith paired difference is equally likely to be $\pm|d_i|$, so the test statistic $\sum_{i=1}^{n} D_i$ is conditionally a sum of independent but not identically distributed binary random variables. The distribution of this sum is the permutation distribution. The situation is identical to that of the preceding section. We can use the uncunx to see whether, given $|d_1|, \ldots, |d_n|$, the permutation distribution is approximately normal.

Most statisticians have a vague notion that when the sample size is large, the permutation distribution is approximately normal, so the permutation test gives the same answer as the t-test asymptotically. But what does this statement really mean? We saw that it is possible to make up values $|d_1|, \ldots, |d_n|$ such that the permutation distribution is not approximately normal, even if n is large. Are such sets of $|d_1|, \ldots, |d_n|$ aberrant, or could they reasonably arise as iid realizations from a distribution with finite variance? To investigate this, click on the uncunx display and press "d" for distribution. This causes the uncunx to choose row deflection sizes iid from a uniform distribution on $[1, 40]$. Despite the fact that this distribution is very spread out and the uncunx only has 9 rows, the permutation distribution is nearly always approximately normal with mean 0 and variance $\sum_{i=1}^n d_i^2$. That is, $Z_n = S_n/(\sum_{i=1}^n d_i^2)^{1/2}$ is approximately standard normal for any values $|d_1|, \ldots, |d_n|$ that might arise as iid observations from this distribution. More generally, for any distribution of row deflection sizes with finite variance, the set of infinite realizations $|d_1|, \ldots, |d_n|, \ldots$ for which the permutation distribution of $Z_n = S_n/(\sum_{i=1}^n d_i^2)^{1/2}$ does not approach $\mathrm{N}(0, 1)$ as $n \to \infty$ has probability 0 of occurring. In other words, we get the same answer asymptotically whether we use a t-test or a permutation test in the paired setting. Van der Vaart (1998) proves the analog of this result in an unpaired setting.

5 The CLT for Symmetric Random Variables

The arguments of the preceding section can be used to deduce that the CLT holds not just for symmetric binary random variables, but for any iid random variables D_i symmetric about 0. The steps consist of: 1) conditioning on $|d_i|,\ i = 1, \ldots, n$, which creates independent, symmetric binary random variables; 2) using the uncunx to conclude that conditioned on the $|d_i|$, S_n is approximately $\mathrm{N}(0, \sum_{i=1}^n d_i^2)$, or equivalently, $Z_n = S_n/(\sum_{i=1}^n d_i^2)^{1/2}$ is approximately standard normal; 3) arguing that because the conditional distribution of Z_n given $|d_i|,\ i = 1, \ldots, n$ is approximately standard normal for all possible values $|d_1|, |d_2| \ldots, |d_n|, \ldots$ (except those in a set of probability 0), Z_n is unconditionally asymptotically standard normal as well; 4) arguing that $\hat{\sigma}^2 = \sum D_i^2/n$ is very close to $\mathrm{E}(D_1)^2 = \sigma^2$, so because $\sum D_i/(\sum D_i^2)^{1/2} = n^{1/2} S_n/\hat{\sigma}$ is approximately standard normal, so is $n^{1/2} S_n/\sigma$. From this argument the student can see that the CLT should hold whenever the iid random variables have a symmetric distribution with mean 0 and finite variance.

6 Random Number Generation

Thus far we have considered random variables symmetric about 0, but an interesting modification of the quincunx allows us to visualize the inverse probability transform for generating random variables with an arbitrary continuous distribution F. In the first row, place the nail at the median of F. The 2 nails in the second row are placed at the first and third quartiles, while the 4 nails in the third row are at percentiles corresponding to probability 1/8, 3/8, 5/8, and 7/8, and similarly for subsequent rows. When a ball strikes the nail in the first row, it deflects either to the left or right, striking either the first or second nail in the second row. The size of the left deflection is not the same as the size of the right deflection. Similarly, if the ball deflects left at the first row, hitting the nail at the 1/4 quantile of the second row, it deflects left or right and strikes either the 1/8 or 3/8 quantile of the third row, etc. (Figure 5). The distribution of balls at the bottom of such a quincunx with a large number of rows will be arbitrarily close to the density f of F.

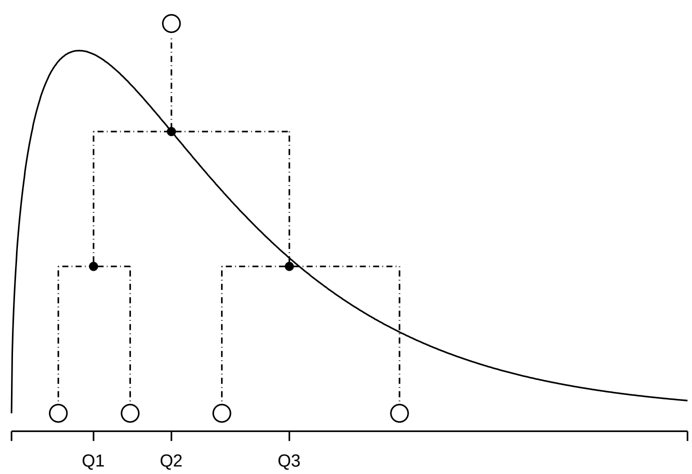

Figure 5: Using a quincunx to simulate from an arbitrary distribution. The nail in the first row divides the distribution into less than or greater than the median. The nail in the second row subdivides into below or above the quartiles Q1 and Q3. As we add rows, the distribution of balls at the bottom approaches that of the continuous curve.

The applet can be used to illustrate this process for the gamma $(2, 1)$ density function $x\exp(-x)$. Click on the display and type "q" for quantile. The first row corresponds to the largest deflection because the distance between the median and the first or third quartile is larger than distances between subsequent quartiles and octiles, etc. The triangles get smaller as we move down the rows. After sufficiently many balls have passed through the uncunx, the histogram strongly resembles the density $f(x)$ (Figure 6).

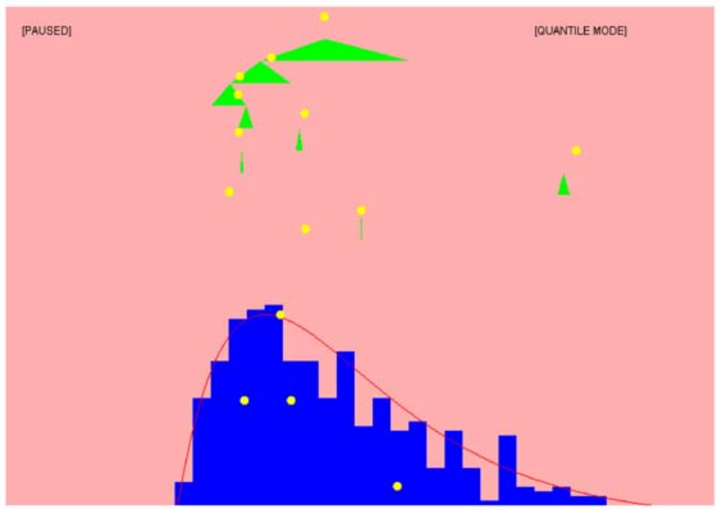

Figure 6: Using the applet to illustrate simulation from an arbitrary distribution with density $f(x)$ using the method described in the legend of Figure 5. The resulting histogram strongly resembles the density $f(x)$.

To see the connection between the applet and the generation of an arbitrary random deviate from a uniform deviate, let $A_i = 0$ or 1 denote whether a ball bounces left or right at row i. Even though the sizes of the deflections from row to row are not independent, the A_i are still iid Bernoulli (1/2) random variables. The quantile mode of the uncunx illustrates the fact that we are able to generate a deviate from F by generating iid Bernoulli (1/2) random variables $A_1, A_2, \ldots A_n \ldots$ Associate with the infinite string $A_1, \ldots A_n, \ldots$ the number in $[0, 1]$ whose base 2 representation is $.A_1 A_2 \ldots, A_n \ldots = A_1/2 + A_2/2^2 + \ldots + A_n/2^n + \ldots = U$. The first digit A_1 tells whether U is in the left half $L = [0, 1/2)$ or right half $R = [1/2, 1)$ of the unit interval; because $\Pr(A_1 = 1) = 1/2$, U is equally likely to be in the left or right half. The second digit A_2 tells whether U is in the left or right half of L or R. Regardless of whether $A_1 = 0$ or 1, $\Pr(A_2 = 1) = 1/2$, so U is equally likely to be in the left or right half of L or R. Together, the first two digits tell us which of the four consecutive intervals $[0, 1/4), [1/4, 1/2), [1/2, 3/4),$ or $[3/4, 1)$ U is in, each being equally likely. This same reasoning can be extended to an arbitrary number of digits. We deduce that, for each n, U is equally likely to be in any of the consecutive intervals of length $1/2^n$. Clearly, U must be uniformly distributed on $[0, 1]$. Hence, if we can generate a uniform deviate U, then its base 2 representation provides an infinite string of left and right branches for the uncunx in quantile mode, which generates a random deviate from distribution F.

Notice that the position of the nail in the first row is $F^{-1}(1/2)$, the positions of the nails in the second row are $F^{-1}(1/4)$ and $F^{-1}(3/4)$, etc. If our uncunx stopped after n rows, the string of n Bernoulli (1/2) left or right deflections is associated with $U_n = A_1/2 + \ldots A_n/2^n$. Note that U_n is within $1/2^n$ of U. The corresponding approximate deviate from F is $F^{-1}(U_n)$, which is very close to $F^{-1}(U)$. In the limit as the number of rows tends to ∞, we are generating a deviate from F by $F^{-1}(U)$. The uncunx therefore gives a visual representation of the well-known result that one can simulate from any continuous distribution function F by $F^{-1}(U)$, where U is uniformly distributed (see, for example, page 156 of DeGroot, 1986).

The uncunx in quantile mode is interesting because the indicators of whether the deflections are left or right are iid Bernoullis, but the deflection sizes are not independent. For example, knowing that the first deflection was to the right pins down the deflection size in the second row to one of two choices; if the first row deflection had instead been to the left, the deflection size choices in the second row would have been different.

7 Summary

The quincunx allowed early statisticians to perform simulations long before the advent of modern computers. The problem is that building even a simple physical quincunx is difficult, and building one that incorporates the modifications in this paper is nearly impossible. Fortunately, the URL (1) provides a virtual universal quincunx that can be used to demonstrate a multitude of useful concepts, including the central limit theorem for independent but not identically distributed random variables, permutation tests in a paired setting, and the generation of a random variable with an arbitrary continuous distribution from a uniform variate. Young students are often enthralled by video games, so the uncunx may provide that spark that sways them to go into the field of statistics. Its versatility makes the uncunx an essential tool in mathematics and statistics classes of all levels.

Acknowledgment

We would like to thank the referees and associate editor for suggestions that improved the paper markedly, and Wendy Lou, who introduced us to each other.

References

1. DeGroot, M.H. (1986). *Probability and Statistics, 2nd ed.* Addison Wesley, Reading Massachusetts.

2. Feller, W. (1957). *An Introduction to Probability Theory and Its Applications*, Vol 1. John Wiley and Sons, New York.

3. Good, P. (2005). *Permutation, Parametric, and Bootstrap Tests of Hypotheses, 3rd ed.* Springer, New York.

4. Kunert, J., Montag, A., and Pohlmann, S. (2001). The quincunx: history and mathematics. *Statistical Papers* **42**, 143-169.

5. Parzen, E. (1960). *Modern Probability Theory and Its Applications.* John Wiley and Sons, New York.

6. Proschan, M.A. (2008). The normal approximation to the binomial. *The American Statistician* **62**, 62-63.

7. Stigler, S.M. (1986). *The History of Statistics: The Measurement of Uncertainty before 1900.* The Belknap Press of Harvard University Press. Cambridge, Massachusetts.

8. van der Vaart, A.W. (1998). Asymptotc Statistics. Cambridge University Press.