THE CENTRAL LIMIT THEOREM AND MIXED DISTRIBUTIONS

Elliot Tanis  Deanna Gross Snyder\textsuperscript{1}  Patricia Lang\textsuperscript{2}

HOPE COLLEGE
Holland, Michigan

Let $X_1, X_2, \ldots, X_n$ be a random sample of size $n$ from a distribution which has mean, $\mu$, and a finite variance, $\sigma^2$.

Let $\overline{X} = (1/n) \sum_{i=1}^{n} X_i$. The Central Limit Theorem states that

$$W_n = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

has a limiting normal distribution with mean, 0, and variance, 1. (Let $N(0,1)$ denote this standard normal distribution.)

When $n$ is "sufficiently large," $W_n$ has an approximate $N(0,1)$ distribution. A "rule of thumb" that is often used is that when $n$ is larger than about 25, $W_n$ will appear to have a $N(0,1)$ distribution. [1]

Note that the hypotheses of the Central Limit Theorem say very little about the distribution from which the sample is taken. In particular the Central Limit Theorem is applicable to mixed distributions.

We shall give two examples which illustrate the conclusion of the Central Limit Theorem empirically when the

\textsuperscript{1} Now at National Security Agency
\textsuperscript{2} Now at the University of Illinois
sample is taken from a mixed distribution. The simulation for these examples was done on Hope College's IBM 1130 computer.

EXAMPLE 1. Roll a fair die. If the outcome is even, let $X$ equal this observed value. If the outcome is odd, spin a spinner which selects a number at random from the interval $[0,1]$. Let $X$ equal the number selected. The random variable, $X$, defined by the "game" has a mixed distribution which has a probability of $1/6$ at each of the integers $2, 4$, and $6$, and has a uniform density of $1/2$ on the interval $[0,1]$.

The distribution function of $X$ is defined by

$$F(x) = \begin{cases} 
0, & x < 0, \\
\frac{x}{2}, & 0 \leq x < 1, \\
\frac{1}{2}, & 1 \leq x < 2, \\
\frac{4}{6}, & 2 \leq x < 4 \\
\frac{5}{6}, & 4 \leq x < 6, \\
1, & 6 \leq x.
\end{cases}$$

The mean and variance of $X$ are defined by

$$\mu = \int_{-\infty}^{\infty} x \, dF(x)$$

$$= \int_{0}^{\infty} \frac{x}{2} \, dx + 2(\frac{1}{6}) + 4(\frac{1}{6}) + 6(\frac{1}{6})$$

$$= \frac{9}{4};$$
\[ \sigma^2 = \int_{-\infty}^{\infty} x^2 \, dF(x) - \left( \frac{9}{4} \right)^2 \]

\[ = \int_{0}^{1} x^2/2 \, dx + 4(1/6) + 16(1/6) + 36(1/6) - (81/16) \]

\[ = 71/16. \]

Let \( X_1, X_2, \ldots, X_n \) denote \( n \) plays of the "game", i.e., a random sample of size \( n \) from the distribution for which \( F \) is the distribution function. For several values of \( n \) we shall use a chi-square goodness of fit test to test the hypothesis, \( H \), that

\[ W_n = \frac{\bar{X} - 9/4}{\sqrt{(71/16)/n}} \]

has a \( N(0,1) \) distribution.

Let the points \(-2.55, -2.25, -1.95, \ldots, 2.25, 2.55\) partition the set of real numbers into 19 intervals. The end intervals are infinite rays and the remaining 17 intervals are of equal length, 0.30. Let \( p_i \) equal the probability that \( W_n \) falls in the \( i \)'th interval when \( H \) is true. E.g., \( p_2 = P(-2.55 < W_n \leq -2.25) \), when \( W_n \) has a \( N(0,1) \) distribution.

We shall generate \( m = 1000 \) \( w_n \)'s on the computer where each \( w_n \) is based on \( n \) plays of the "game." Let \( y_i \) denote the observed number of \( w_n \)'s in the \( i \)'th interval. Then

\[ \chi^2 = \sum_{i=1}^{19} \frac{(y_i - 1000p_i)^2}{1000p_i} \]

has an approximate chi-square distribution with \( \nu = 18 \).
degrees of freedom when $H$ is true.

Instead of setting a significance level for this test, for each value of $n$ between 2 and 15, inclusive, we ran three independent chi-square goodness of fit tests, each based on a sample of size $m = 1000$ $W_n$'s. In table 1 we have recorded the calculated values of $\chi^2$. For this chi-square value we have also recorded the $(1-p)'$th quantile, $\chi^2(p)$. I.e., if $U$ has a chi-square distribution with $v$ degrees of freedom, then

$$P(U \geq \chi^2(p)) = p.$$ 

When $\epsilon$ is recorded, $\chi^2(p) < .0005$.

For this mixed distribution, $W_n$ appears to have a $N(0,1)$ distribution when $n$ is larger than about 8 or 9.

In order to show graphically the convergence of $W_n$ to the $N(0,1)$ distribution, in figures 1, 2, 3, and 4 we have graphed the probability density function for the $N(0,1)$ distribution along with the relative frequency histogram for each set of data which yielded the starred chi-square values in table 1.

EXAMPLE 2. Toss an unbiased coin. If the coin lands heads, let $X = 2$. If the coin lands tails, spin a spinner which selects a number at random from the interval $[0,1]$. Let $X$ equal the number selected.

The random variable, $X$, defined by this game has a mixed distribution. The distribution function of $X$ is defined by
\[
G(x) = \begin{cases} 
0, & x < 0, \\
\frac{x}{2}, & 0 \leq x < 1, \\
\frac{1}{2}, & 1 \leq x < 2, \\
1, & 2 \leq x.
\end{cases}
\]

The mean and variance of \( X \) are defined by

\[
\mu = \int_{-\infty}^{\infty} x \, dF(x)
\]

\[
= \int_{0}^{\infty} \frac{x}{2} \, dx + 2(1/2)
\]

\[
= \frac{5}{4};
\]

\[
\sigma^2 = \int_{-\infty}^{\infty} x^2 \, dF(x) - (\frac{5}{4})^2
\]

\[
= \int_{0}^{\infty} \frac{x^2}{2} \, dx + 4(1/2) - (25/16)
\]

\[
= \frac{29}{48}.
\]

For \( n \) between 2 and 15, inclusive, we shall again use a chi-square goodness of fit test to test the hypothesis, \( H \), that

\[
W_n = \frac{\bar{X} - \frac{5}{4}}{\sqrt{\frac{29}{48}/n}}
\]

has a \( N(0,1) \) distribution.

For this example we shall partition the set of real numbers into 20 intervals, each of which has a probability of .05 when \( H \) is true.

Let \( Y_i \) equal the observed number of \( W_n \)'s in the \( i \)'th interval. If \( m = 1000 \) \( W_n \)'s are observed,
\[ x^2 = \sum_{i=1}^{20} \frac{(Y_i - 50)^2}{50} \]

has an approximate chi-square distribution with \( v = 19 \) degrees of freedom when \( H \) is true.

In table 2 we have recorded the calculated values of \( x^2 \) along with \( x^2(p) \) for each of three independent trials. As in table 1, \( \varepsilon \) is recorded when \( x^2(p) < .0005 \). From table 2 we see that \( W_n \) appears to have a \( N(0,1) \) distribution when \( n \) is 8 or larger.

In figures 5, 6, 7, 8, and 9 we have graphed the probability density function for the \( N(0,1) \) distribution along with the relative frequency histogram for each set of data which yielded the starred chi-square values in table 2.
<table>
<thead>
<tr>
<th>n</th>
<th>$\chi^2$</th>
<th>$\chi^2(p)$</th>
<th>$\chi^2$</th>
<th>$\chi^2(p)$</th>
<th>$\chi^2$</th>
<th>$\chi^2(p)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>753.20</td>
<td>$\varepsilon$</td>
<td>639.07</td>
<td>$\varepsilon$</td>
<td>614.34</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>3</td>
<td>247.60*</td>
<td>$\varepsilon$</td>
<td>160.22</td>
<td>$\varepsilon$</td>
<td>189.49</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>4</td>
<td>76.72</td>
<td>$\varepsilon$</td>
<td>100.44</td>
<td>$\varepsilon$</td>
<td>106.72</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>5</td>
<td>48.76</td>
<td>$\varepsilon$</td>
<td>52.85</td>
<td>$\varepsilon$</td>
<td>48.61</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>6</td>
<td>23.91*</td>
<td>.158</td>
<td>48.17</td>
<td>$\varepsilon$</td>
<td>35.18</td>
<td>.009</td>
</tr>
<tr>
<td>7</td>
<td>30.50</td>
<td>.033</td>
<td>32.22</td>
<td>.021</td>
<td>47.76</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>8</td>
<td>26.11</td>
<td>.097</td>
<td>29.66</td>
<td>.041</td>
<td>24.09</td>
<td>.152</td>
</tr>
<tr>
<td>9</td>
<td>11.70</td>
<td>.862</td>
<td>27.98</td>
<td>.062</td>
<td>20.09*</td>
<td>.328</td>
</tr>
<tr>
<td>10</td>
<td>15.48</td>
<td>.629</td>
<td>22.43</td>
<td>.213</td>
<td>18.11</td>
<td>.449</td>
</tr>
<tr>
<td>11</td>
<td>25.43</td>
<td>.114</td>
<td>29.68</td>
<td>.041</td>
<td>25.53</td>
<td>.111</td>
</tr>
<tr>
<td>12</td>
<td>17.65</td>
<td>.479</td>
<td>13.23*</td>
<td>.778</td>
<td>27.96</td>
<td>.063</td>
</tr>
<tr>
<td>13</td>
<td>22.75</td>
<td>.200</td>
<td>13.01</td>
<td>.791</td>
<td>13.83</td>
<td>.740</td>
</tr>
<tr>
<td>14</td>
<td>12.06</td>
<td>.844</td>
<td>23.25</td>
<td>.181</td>
<td>34.02</td>
<td>.013</td>
</tr>
</tbody>
</table>

Table 1

* See figures 1, 2, 3, and 4.
$N = 3, \text{ CHI-SQUARE} = 247.60$

**FIGURE 1**
$N = 6$, CHI-SQUARE = 23.91

FIGURE 2
\[ N = 9, \text{ CHI-SQUARE} = 20.09 \]

**FIGURE 3**
<table>
<thead>
<tr>
<th>n</th>
<th>$\chi^2$</th>
<th>$\chi^2(p)$</th>
<th>$\chi^2$</th>
<th>$\chi^2(p)$</th>
<th>$\chi^2$</th>
<th>$\chi^2(p)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1222.20</td>
<td>ε</td>
<td>1178.80</td>
<td>ε</td>
<td>1269.96</td>
<td>ε</td>
</tr>
<tr>
<td>3</td>
<td>345.92</td>
<td>ε</td>
<td>382.24</td>
<td>ε</td>
<td>399.44*</td>
<td>ε</td>
</tr>
<tr>
<td>4</td>
<td>202.04</td>
<td>ε</td>
<td>184.44</td>
<td>ε</td>
<td>224.56</td>
<td>ε</td>
</tr>
<tr>
<td>5</td>
<td>104.96</td>
<td>ε</td>
<td>110.60</td>
<td>ε</td>
<td>102.12</td>
<td>ε</td>
</tr>
<tr>
<td>6</td>
<td>38.60*</td>
<td>.005</td>
<td>46.56</td>
<td>ε</td>
<td>41.40</td>
<td>.002</td>
</tr>
<tr>
<td>7</td>
<td>30.24</td>
<td>.049</td>
<td>37.20</td>
<td>.007</td>
<td>28.84</td>
<td>.069</td>
</tr>
<tr>
<td>8</td>
<td>22.64</td>
<td>.254</td>
<td>27.20</td>
<td>.100</td>
<td>17.56</td>
<td>.552</td>
</tr>
<tr>
<td>9</td>
<td>22.40</td>
<td>.265</td>
<td>20.60*</td>
<td>.359</td>
<td>15.28</td>
<td>.705</td>
</tr>
<tr>
<td>10</td>
<td>18.08</td>
<td>.517</td>
<td>22.48</td>
<td>.261</td>
<td>12.84</td>
<td>.847</td>
</tr>
<tr>
<td>11</td>
<td>38.44</td>
<td>.005</td>
<td>6.20</td>
<td>.997</td>
<td>12.40</td>
<td>.868</td>
</tr>
<tr>
<td>12</td>
<td>16.00*</td>
<td>.657</td>
<td>31.52</td>
<td>.035</td>
<td>16.24</td>
<td>.641</td>
</tr>
<tr>
<td>13</td>
<td>14.88</td>
<td>.730</td>
<td>18.92</td>
<td>.462</td>
<td>25.40</td>
<td>.148</td>
</tr>
<tr>
<td>14</td>
<td>21.96</td>
<td>.286</td>
<td>24.16</td>
<td>.190</td>
<td>15.92</td>
<td>.663</td>
</tr>
<tr>
<td>15</td>
<td>23.80</td>
<td>.204</td>
<td>12.64*</td>
<td>.856</td>
<td>21.40</td>
<td>.315</td>
</tr>
</tbody>
</table>

Table 2

* See figures 5, 6, 7, 8, and 9.
N = 3, CHI-SQUARE = 339.44

FIGURE 5
$N = 6$, CHI-SQUARE $= 38.60$

FIGURE 6
N = 9, CHI-SQUARE = 20.60

FIGURE 7
N = 12, CHI-SQUARE = 15.00

FIGURE B
REFERENCES