The Modern Form of Bernoulli’s Theorem

The modern form of Bernoulli’s theorem is:

If N is sufficiently large, the probability that \( \left| \frac{M}{N} - P \right| < \varepsilon \) will be greater than \( 1 - \eta \).

M is the number of successes in N trials, \( P \) is the probability of a success in a single trial and \( \varepsilon \) and \( \eta \) are positive numbers chosen as small as desired.

In this paper, I outline the proof of this theorem using Bernoulli’s and Markov’s techniques.

One of the improvements that Markov made to Bernoulli’s theorem was to remove the restrictions that Bernoulli put on N and \( \varepsilon \). In Bernoulli’s proof \( \varepsilon \) had to be \( \frac{1}{T} \) and N had to be a multiple of T. Bernoulli used \( NT \) to stand for the number of trials. I will use \( N^* \) to stand for Bernoulli’s use of N and N will stand for the number of trials.

\( NP \) roughly corresponds to \( N^*R \)
\( NP + N \varepsilon \) roughly corresponds to \( N^*R + N^* \)
\( NP - N \varepsilon \) roughly corresponds to \( N^*R - N^* \)

I say roughly because \( NP \), \( NP + N \varepsilon \), and \( NP - N \varepsilon \) may not be integers.
Let $\lambda$ be the smallest integer greater than or equal to $NP$.
Let $\mu$ be the smallest integer greater than or equal to $NP + N \varepsilon$.
Let $k$ be the largest integer less than or equal to $NP - N \varepsilon$.
Let $T_i$ be the probability of getting exactly $i$ successes in $N$ trials.

The probability that $\left| \frac{M}{N} - P \right| < \varepsilon$ is equal to:

$$T_{k+1}+T_{k+2}+ \ldots + T_\lambda + T_{\lambda+1}+ \ldots + T_\mu + T_{\mu+1}+ \ldots + T_N$$

Let $A = T_\lambda + T_{\lambda+1}+ \ldots + T_\mu$ and $B = T_{\mu+1}+ \ldots + T_N$

Let $C = T_\mu + T_{\mu+1}+ \ldots + T_N$ and $D = T_k + T_{k-1}+ \ldots + T_0$

The probability that $\left| \frac{M}{N} - P \right| < \varepsilon$ is $A + B$ and $A+B+C+D = 1$.

We will show that if $N$ is sufficiently large, $C< \frac{A \eta}{1-\eta}$ and $D< \frac{B \eta}{1-\eta}$

So then $A+B + \frac{A \eta}{1-\eta} + \frac{B \eta}{1-\eta} > 1$. So $A+B > 1 - \eta$.

So if $N$ is sufficiently large, the probability that $\left| \frac{M}{N} - P \right| < \varepsilon$ is greater than $1 - \eta$. 

2
\[
\begin{align*}
\frac{T_{\mu}}{T_{\lambda}} & > \frac{T_{\mu+1}}{T_{\lambda+1}} > \frac{T_{\mu+2}}{T_{\lambda+2}} > \ldots > \frac{T_{N}}{T_{N-(\mu-\lambda)}} \\
\text{Let } A_1 &= T_{\mu} + T_{\mu+1} + \ldots + T_{2\mu-\lambda-1} \\
A_2 &= T_{2\mu-\lambda} + T_{2\mu-\lambda+1} + \ldots + T_{3\mu-2\lambda-1} \\
A_3 &= T_{3\mu-2\lambda} + T_{3\mu-2\lambda+1} + \ldots + T_{4\mu-3\lambda-1} \\
&\quad \vdots \\
\frac{T_{\mu}}{T_{\lambda}} &> \frac{A_1}{A}, \quad \frac{T_{\mu}}{T_{\lambda}} > \frac{A_2}{A}, \quad \frac{T_{\mu}}{T_{\lambda}} > \frac{A_3}{A}, \quad \frac{T_{\mu}}{T_{\lambda}} > \frac{A_4}{A}, \quad \ldots \\
\text{So } \left(\frac{T_{\mu}}{T_{\lambda}}\right)^2 &> \frac{A_2}{A}, \quad \left(\frac{T_{\mu}}{T_{\lambda}}\right)^3 > \frac{A_3}{A}, \quad \left(\frac{T_{\mu}}{T_{\lambda}}\right)^4 > \frac{A_4}{A}, \quad \ldots \\
\text{So } \frac{T_{\mu}}{T_{\lambda}} + \left(\frac{T_{\mu}}{T_{\lambda}}\right)^2 + \left(\frac{T_{\mu}}{T_{\lambda}}\right)^3 + \left(\frac{T_{\mu}}{T_{\lambda}}\right)^4 + \ldots &> \frac{A_1 + A_2 + A_3 + \ldots}{A} \\
\text{So } \frac{T_{\mu}}{T_{\lambda}} + \left(\frac{T_{\mu}}{T_{\lambda}}\right)^2 + \left(\frac{T_{\mu}}{T_{\lambda}}\right)^3 + \left(\frac{T_{\mu}}{T_{\lambda}}\right)^4 + \ldots &> \frac{T_{\mu} + T_{\mu+1} + \ldots + T_N}{A} \\
\text{So } \frac{T_{\mu}}{1 - \frac{T_{\mu}}{T_{\lambda}}} &> \frac{T_{\mu} + T_{\mu+1} + \ldots + T_N}{A} = \frac{C}{A}. \quad \text{So } A \left(\frac{T_{\mu}}{1 - \frac{T_{\mu}}{T_{\lambda}}}\right) > C
\end{align*}
\]
- Since \( k < \lambda - 1 \), \( \frac{T_k}{T_{\lambda - 1}} > \frac{T_{k-1}}{T_{\lambda - 2}} > \frac{T_{k-2}}{T_{\lambda - 3}} > \ldots > \frac{T_0}{T_{\lambda - k - 1}} \)

- Let \( B_1 = T_k + T_{k-1} + \ldots + T_{2k-\lambda + 2} \)

\[
B_2 = T_{2k-\lambda + 1} + T_{2k-\lambda} + \ldots + T_{3k-2\lambda + 3}
\]

\[
B_3 = T_{3k-2\lambda + 2} + T_{3k-2\lambda + 1} + \ldots + T_{4k-3\lambda + 4}
\]

\[
\begin{align*}
\frac{T_k}{T_{\lambda - 1}} &> \frac{B_1}{B} , \quad \frac{T_k}{T_{\lambda - 1}} > \frac{B_2}{B_1} , \quad \frac{T_k}{T_{\lambda - 1}} > \frac{B_3}{B_2} \ldots \\
\frac{T_k}{T_{\lambda - 1}} &> \frac{B_1}{B} , \quad \left(\frac{T_k}{T_{\lambda - 1}}\right)^2 > \frac{B_2}{B} , \quad \left(\frac{T_k}{T_{\lambda - 1}}\right)^3 > \frac{B_3}{B} \ldots \\
\text{So} \quad \frac{T_k}{T_{\lambda - 1}} + \left(\frac{T_k}{T_{\lambda - 1}}\right)^2 + \left(\frac{T_k}{T_{\lambda - 1}}\right)^3 + \ldots > \frac{B_1 + B_2 + B_3 + \ldots}{B} \\
\text{So} \quad \frac{T_k}{T_{\lambda - 1}} + \left(\frac{T_k}{T_{\lambda - 1}}\right)^2 + \left(\frac{T_k}{T_{\lambda - 1}}\right)^3 + \ldots > \frac{T_k + T_{k-1} + \ldots + T_0}{B} = \frac{D}{B}
\end{align*}
\]
So $\frac{T_k}{T_{k-1}} > \frac{D}{B}$, So $B\frac{T_k}{T_{k-1}} > D$

- It remains to be shown that by making $N$ sufficiently large

$\frac{T_\mu}{T_\lambda}$ and $\frac{T_k}{T_{k-1}}$ can be made smaller than $\eta$.

We will show that $\frac{T_\lambda}{T_\mu}$ and $\frac{T_{\lambda-1}}{T_k}$ can be made as large as desired by making $N$ sufficiently large, which gives the same result.

We will apply my method from Lemma 7 of my paper Bernoulli’s Theorem.

We calculate $\frac{T_\lambda}{T_\mu}$ from the formula 

$$P(K) = \frac{N! P^K Q^{N-K}}{K!(N-K)!}$$

Where $P(K)$ is the probability of getting exactly $K$ successes in $N$ trials when the probability of success on a single trial is $P$ and $Q = 1 - P$.

$$T_\lambda = \frac{N(N-1)....(N-\lambda+1)P^\lambda Q^{N-\lambda}}{\lambda(\lambda-1)......1}$$
\[ T_\mu = \frac{N(N - 1)(N - \mu + 1)P^\mu Q^{N-\mu}}{\mu(\mu - 1)\ldots 1} \]

So \[ \frac{T_\lambda}{T_\mu} = \frac{\mu(\mu - 1)(\lambda + 1)Q^{\mu - \lambda}}{(N - \lambda)(N - \lambda - 1)(N - \mu + 1)P^{\mu - \lambda}} \]

Reversing the order of the factors in both the numerator and denominator so they will be increasing from left to right instead of decreasing we get:

\[ \frac{T_\lambda}{T_\mu} = \frac{(\lambda + 1)(\lambda + 2)(\mu - 1)Q^{\mu - \lambda}}{(N - \mu + 1)(N - \mu + 2)\ldots (N - \lambda - 1)(N - \lambda)P^{\mu - \lambda}} \]

\[ \frac{T_\lambda}{T_\mu} = \frac{(\lambda Q + Q)}{(NP - \mu P + P)} \ast \frac{(\lambda Q + 2Q)}{(NP - \mu P + 2P)} \ast \ldots \ast \frac{(\mu Q - Q)}{(NP - \lambda P - P)} \ast \frac{\mu Q}{(NP - \lambda P)} \]

Notice that each fraction is obtained from the previous fraction by adding \( Q \) to the numerator and \( P \) to the denominator.

The first fraction is itself obtained from \( \frac{\lambda Q}{NP - \mu P} \) by adding \( Q \) to the numerator and \( P \) to the denominator. There are \( \mu - \lambda \) fractions in the product, so by the same reasoning as in lemma 7, \( \frac{T_\lambda}{T_\mu} \) is
greater than the smaller of \( \left( \frac{\lambda Q}{NP - \mu P} \right)^{\mu - \lambda} \) or \( \left( \frac{\mu Q}{NP - \lambda P} \right)^{\mu - \lambda} \).

If \( NP \) and \( N\epsilon \) are both integers then \( \mu - \lambda = N\epsilon \). If one or both are not integers, then \( \mu - \lambda > N\epsilon - 1 \).

So \( \frac{T_\lambda}{T_\mu} \) is greater than the smaller of \( \left( \frac{\lambda Q}{NP - \mu P} \right)^{N\epsilon - 1} \) or \( \left( \frac{\mu Q}{NP - \lambda P} \right)^{N\epsilon - 1} \).

\[
\frac{(NP + N\epsilon)Q}{NP - NP^2} \leq \frac{\mu Q}{NP - \lambda P} \quad \text{and} \quad \frac{(NP + N\epsilon)Q}{NP - NP^2} = \frac{P + \epsilon}{P}.
\]

\[
\frac{NPQ}{NP - (NP + N\epsilon)P} \leq \frac{\lambda Q}{NP - \mu P} \quad \text{and} \quad \frac{NPQ}{NP - (NP + N\epsilon)P} = \frac{Q}{Q - \epsilon}.
\]

So \( \frac{T_\lambda}{T_\mu} \) is greater than the smaller of \( \left( \frac{P + \epsilon}{P} \right)^{N\epsilon - 1} \) or \( \left( \frac{Q}{Q - \epsilon} \right)^{N\epsilon - 1} \).

So by making \( N \) sufficiently large, \( \frac{T_\lambda}{T_\mu} \) can be made greater than \( \frac{1}{\eta} \).

Using \( P(K) = \frac{N!P^K Q^{N-K}}{K!(N-K)!} \)

\[
T_{\lambda - 1} = \frac{N(N-1)\ldots(N-\lambda + 2)P^{\lambda - 1}Q^{N-\lambda + 1}}{(\lambda - 1)(\lambda - 2)\ldots 1}
\]

\[
T_k = \frac{N(N-1)\ldots(N-k + 1)P^k Q^{N-k}}{k(k-1)\ldots 1}
\]
So \[ \frac{T_{\lambda-1}}{T_k} = \frac{(N - k)(N - k - 1)\ldots(N - \lambda + 2)P^{\lambda-k-1}}{(\lambda - 1)(\lambda - 2)\ldots(k + 1)Q^{\lambda-k-1}} \]

Reversing the order of the factors in both the numerator and denominator, so that the factors will be increasing instead of decreasing gives:

\[ \frac{T_{\lambda-1}}{T_k} = \frac{N - \lambda + 2}{k + 1} \cdot \frac{N - \lambda + 3}{k + 2} \cdot \ldots \cdot \frac{N - k}{\lambda - 1} \cdot \frac{P^{\lambda-k-1}}{Q^{\lambda-k-1}} \]

So \[ \frac{T_{\lambda-1}}{T_k} = \frac{NP - \lambda P + 2P}{kQ + Q} \cdot \frac{NP - \lambda P + 3P}{kQ + 2Q} \cdot \ldots \cdot \frac{NP - kP}{\lambda Q - Q} \]

Notice that each fraction is obtained from the previous fraction by adding \( P \) to the numerator and \( Q \) to the denominator. The first fraction is itself obtained from \( \frac{NP - \lambda P + P}{kQ} \) by adding \( P \) to the numerator and \( Q \) to the denominator. There are \( \lambda - k - 1 \) fractions in the product, so by the same reasoning as in lemma 7,

\[ \frac{T_{\lambda-1}}{T_k} \] is greater than the smaller of \( \left( \frac{NP - \lambda P + P}{kQ} \right)^{\lambda-k-1} \) or \( \left( \frac{NP - kP}{\lambda Q - Q} \right)^{\lambda-k-1} \).

If \( NP \) and \( Ne \) are both integers then \( \lambda - k - 1 = Ne - 1 \) and this is the smallest value that \( \lambda - k - 1 \) can have.

So \[ \frac{T_{\lambda-1}}{T_k} \] is greater than the smaller of \( \left( \frac{NP - \lambda P + P}{kQ} \right)^{Ne-1} \) or \( \left( \frac{NP - kP}{\lambda Q - Q} \right)^{Ne-1} \)

\[ \frac{NP - \lambda P + P}{kQ} > \frac{NP - (NP + 1)P + P}{(NP - Ne)Q} = \frac{NP(1 - P)}{(NP - Ne)(1 - P)} = \frac{P}{P - \epsilon} \]
\[
\frac{NP - kP}{\lambda Q - Q} > \frac{NP - (NP - N\varepsilon)P}{(NP + 1)Q - Q} = \frac{NP(1 - P) + N\varepsilon P}{NPQ} = \frac{Q + \varepsilon}{Q}.
\]

So \( \frac{T_{\lambda-1}}{T_k} \) is greater than the smaller of \( \left( \frac{P}{P - \varepsilon} \right)^{Ne-1} \) or \( \left( \frac{Q + \varepsilon}{Q} \right)^{Ne-1} \).

So by making \( N \) sufficiently large, \( \frac{T_{\lambda-1}}{T_k} \) can be made greater than \( \frac{1}{\eta} \).

So if \( N \) is sufficiently large \( \frac{T_{\lambda}}{T_\mu} > \frac{1}{\eta} \) and \( \frac{T_{\lambda-1}}{T_k} > \frac{1}{\eta} \).

So \( \frac{T_\mu}{T_{\lambda}} < \eta \) and \( \frac{T_k}{T_{\lambda-1}} < \eta \).

So if \( N \) is sufficiently large, \( C < A \left( \frac{\eta}{1-\eta} \right) \) and \( D < B \cdot \frac{\eta}{1-\eta} \).

This completes the proof.

Daniel Daniels