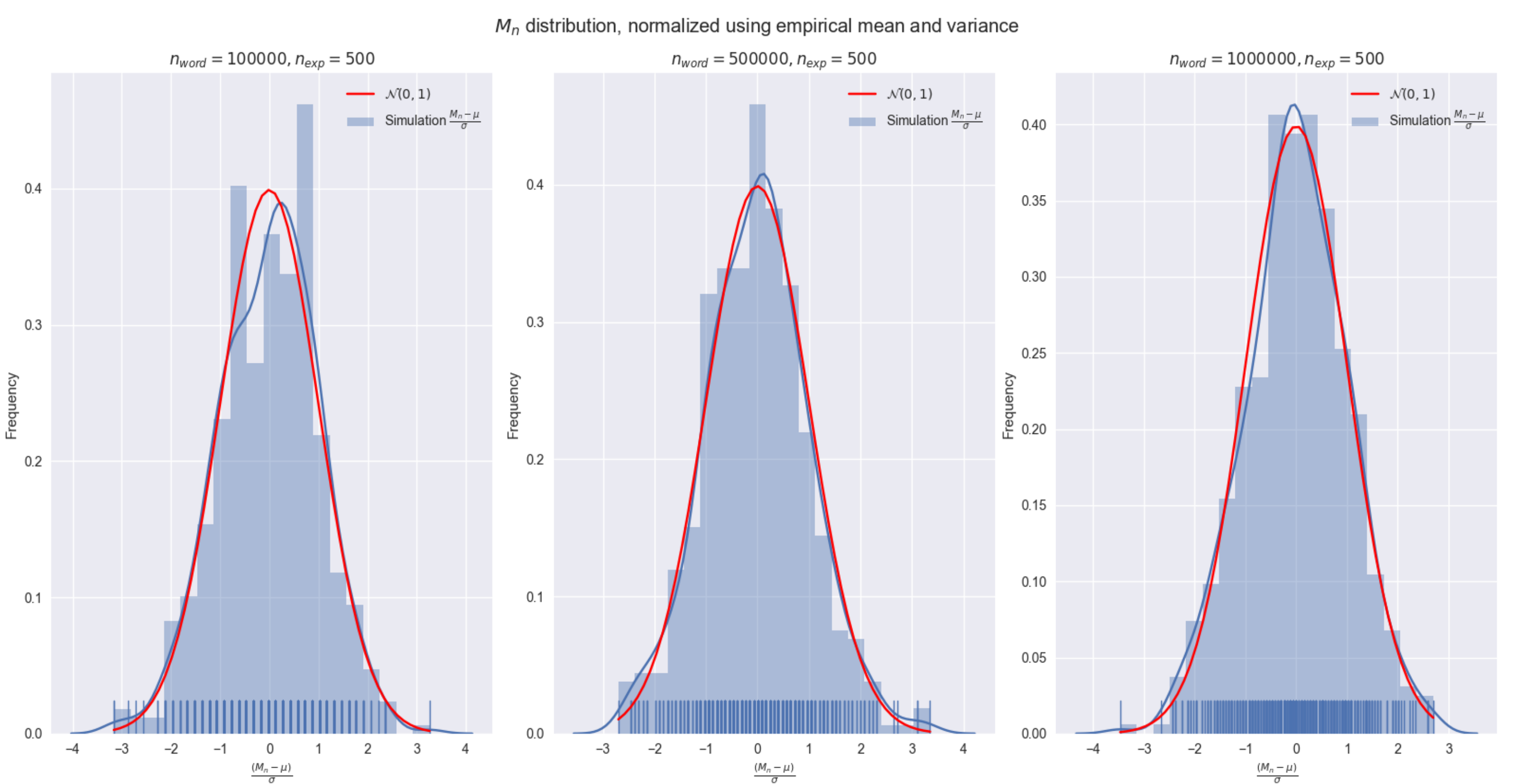
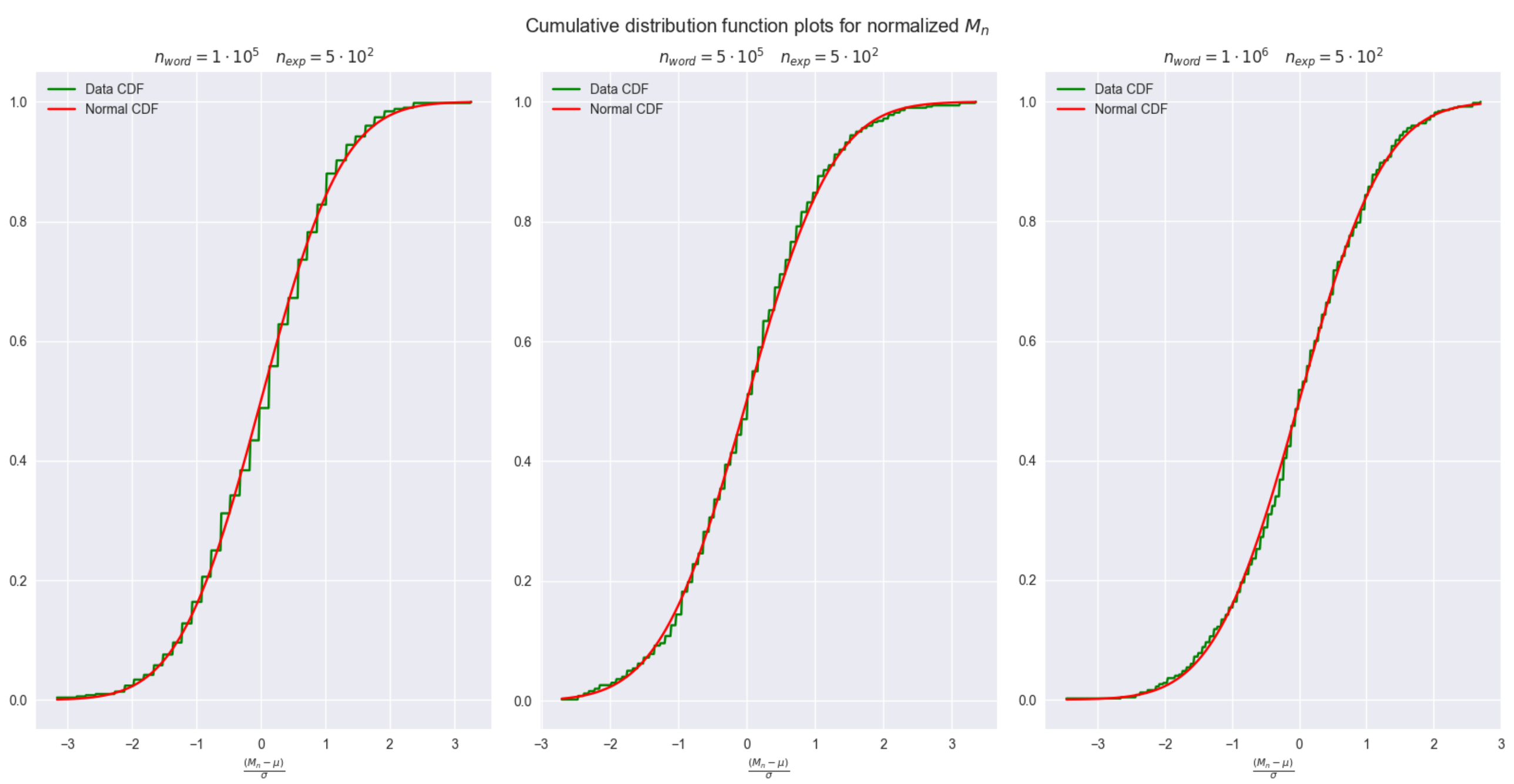
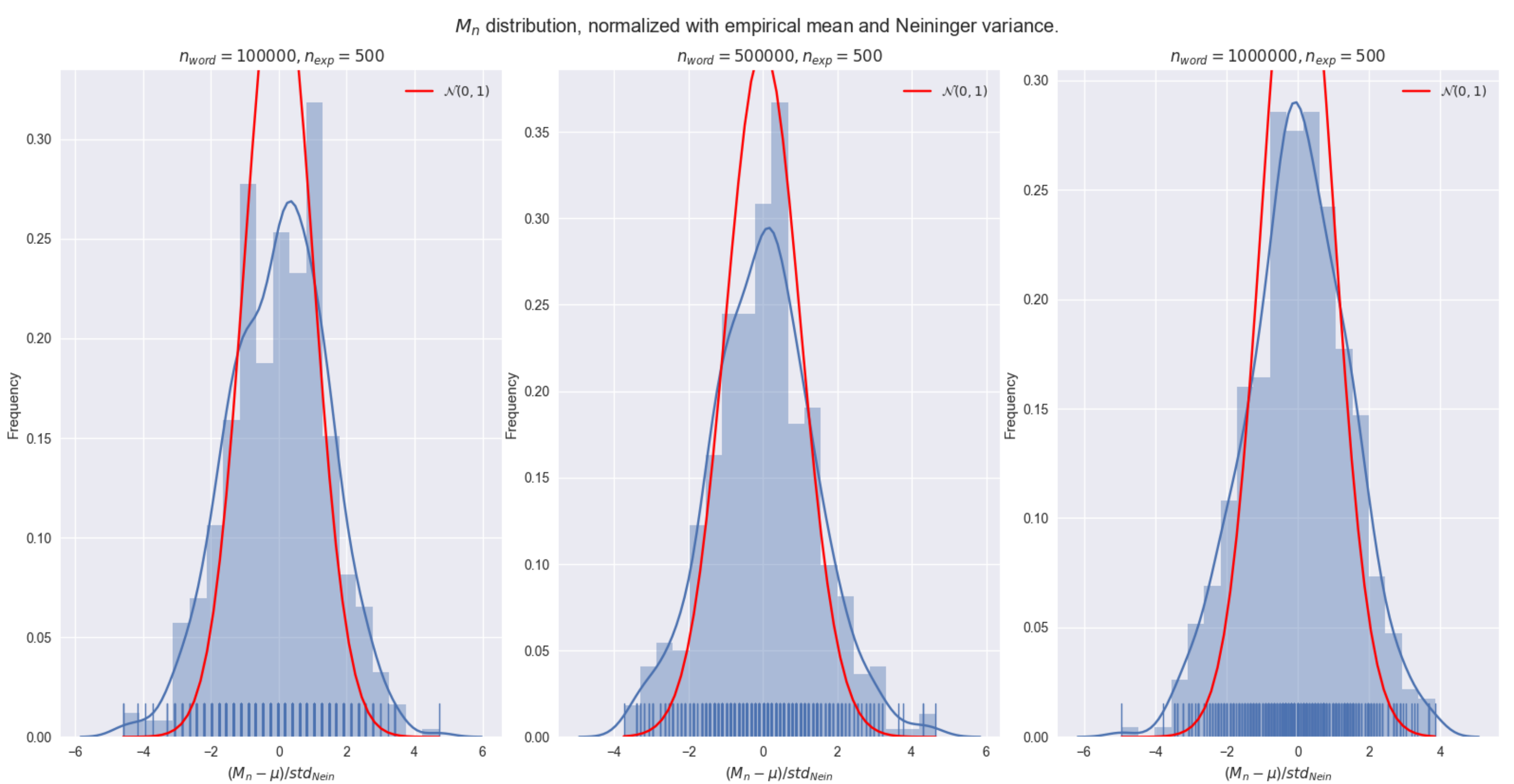
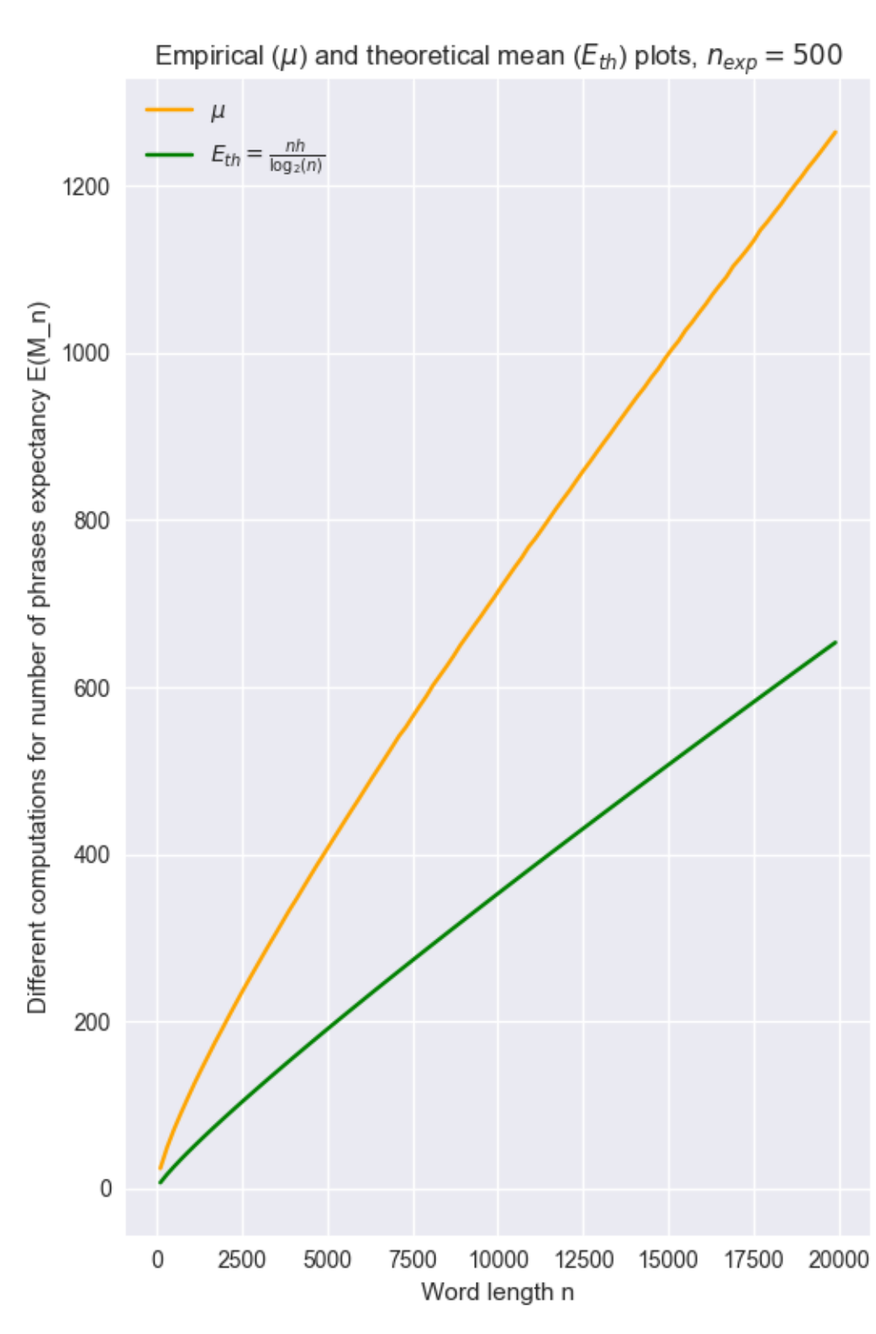
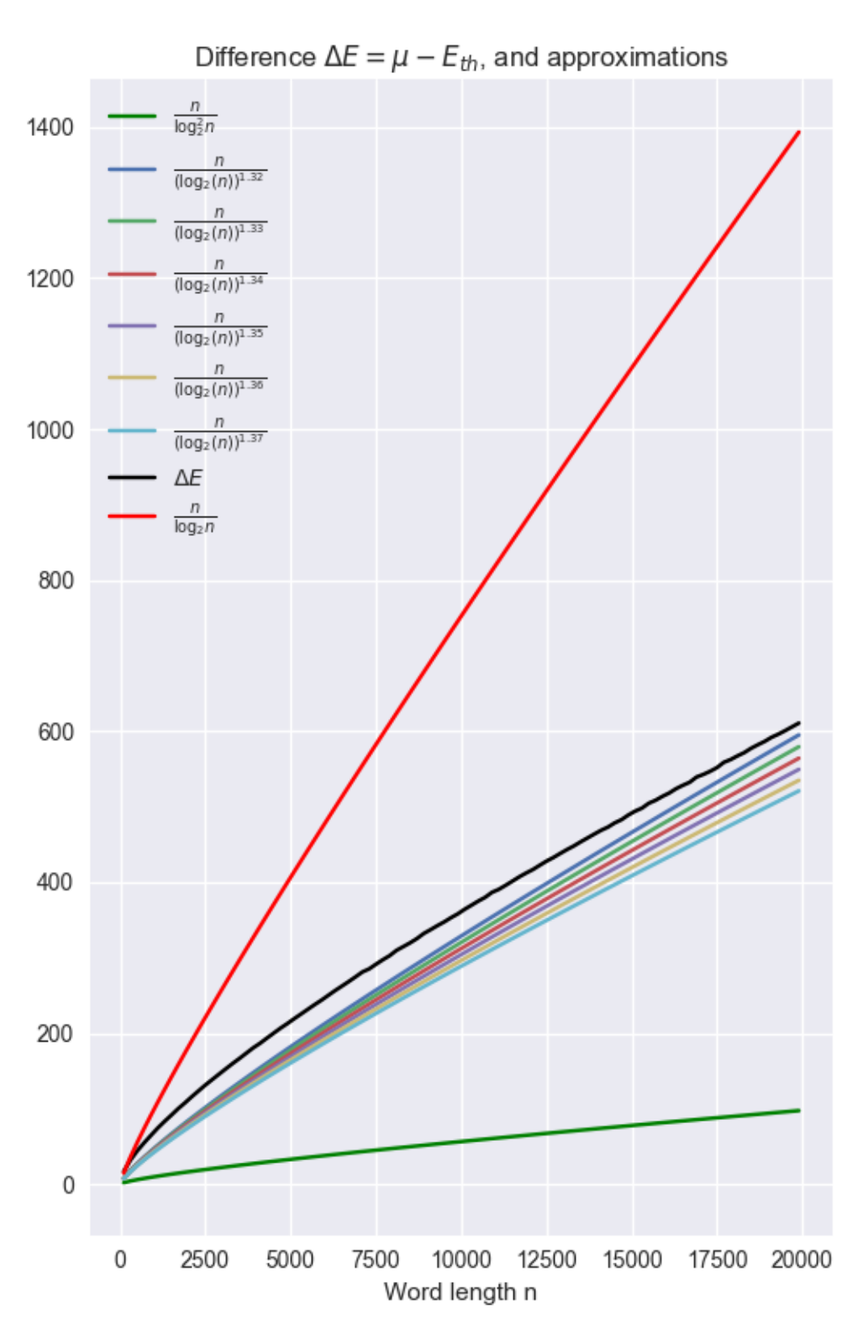

Histograms of the values of M_n for different word lengths
$n_{word} = 100000, n_{exp} = 500$
$n_{word} = 500000, n_{exp} = 500$
$n_{word} = 1000000, n_{exp} = 500$
Counts
M_n
M_n
M_n

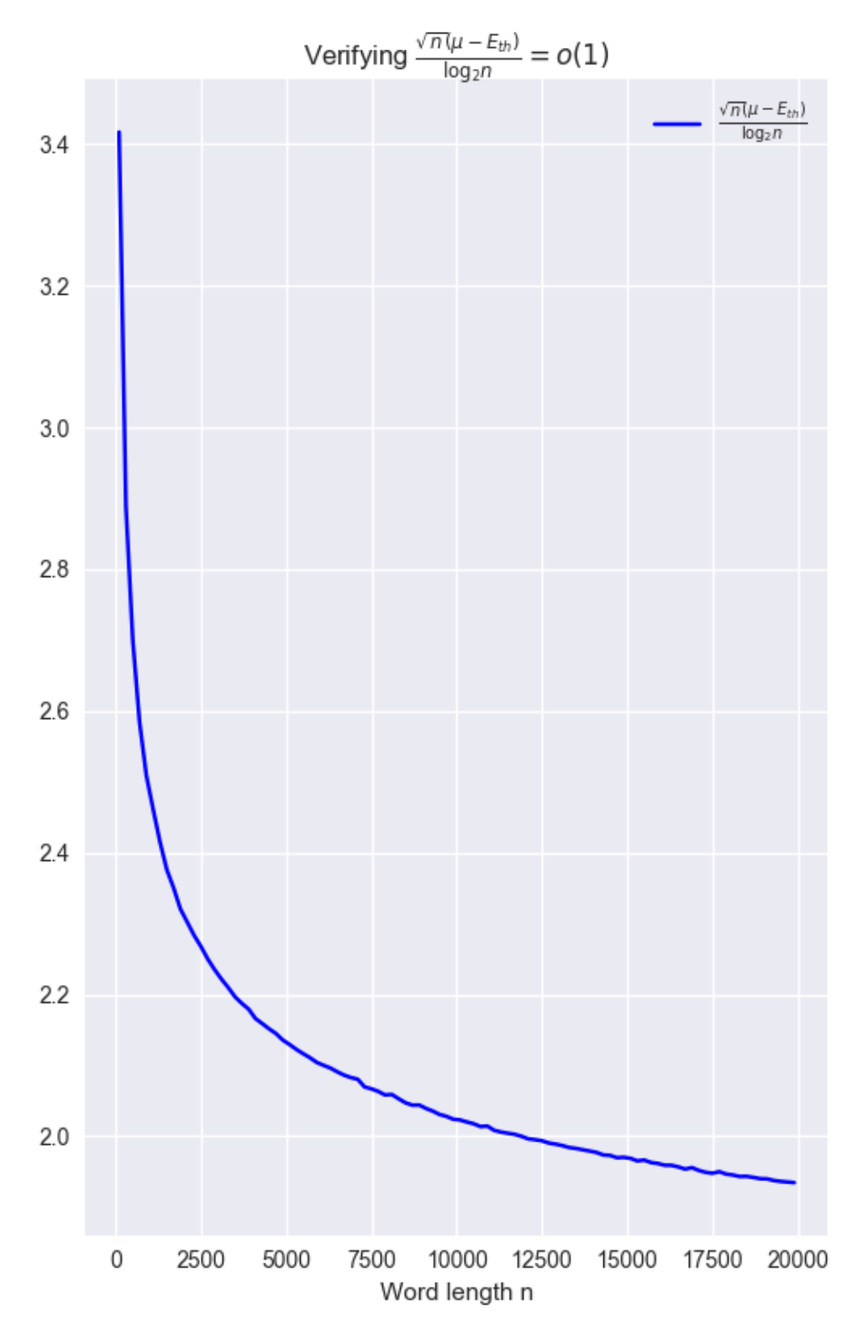

M_n distribution, normalized using empirical mean and variance
$n_{word} = 100000, n_{exp} = 500$
$n_{word} = 500000, n_{exp} = 500$
$n_{word} = 1000000, n_{exp} = 500$
$\mathcal{N}(0, 1)$
Simulation $\frac{M_n - \mu}{\sigma}$
Frequency
$\frac{(M_n - \mu)}{\sigma}$

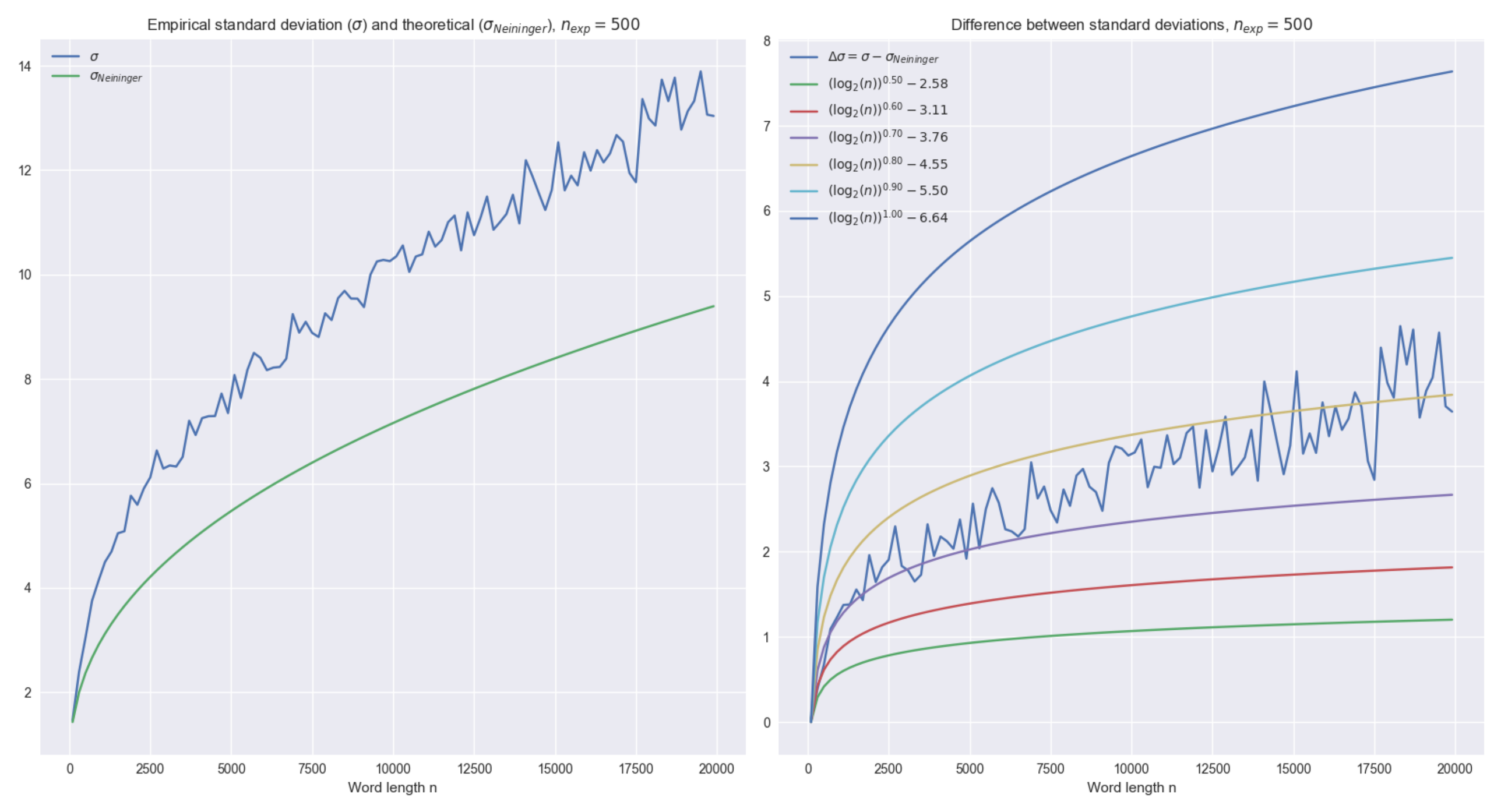

Cumulative distribution function plots for normalized M_n
$n_{word} = 1 \cdot 10^5 \quad n_{exp} = 5 \cdot 10^2$
$n_{word} = 5 \cdot 10^5 \quad n_{exp} = 5 \cdot 10^2$
$n_{word} = 1 \cdot 10^6 \quad n_{exp} = 5 \cdot 10^2$
Data CDF
Normal CDF
0.0
0.2
0.4
0.6
0.8
1.0
−3
−2
−1
0
1
2
3
$\frac{(M_n - \mu)}{\sigma}$


M_n distribution, normalized with empirical mean and Neininger variance.
$n_{word} = 100000, n_{exp} = 500$
$n_{word} = 500000, n_{exp} = 500$
$n_{word} = 1000000, n_{exp} = 500$
$\mathcal{N}(0, 1)$
Frequency
$(M_n - \mu)/std_{Nein}$


Empirical (μ) and theoretical mean (E_{th}) plots, $n_{exp} = 500$
μ
$E_{th} = \frac{nh}{\log_2(n)}$
Different computations for number of phrases expectancy E(M_n)
Word length n


Difference $\Delta E = \mu - E_{th}$, and approximations
$\frac{n}{\log_2^2 n}$
$\frac{n}{(\log_2(n))^{1.32}}$
$\frac{n}{(\log_2(n))^{1.33}}$
$\frac{n}{(\log_2(n))^{1.34}}$
$\frac{n}{(\log_2(n))^{1.35}}$
$\frac{n}{(\log_2(n))^{1.36}}$
$\frac{n}{(\log_2(n))^{1.37}}$
ΔE
$\frac{n}{\log_2 n}$
Word length n


Verifying $\frac{\sqrt{n}(\mu - E_{th})}{\log_2 n} = o(1)$
$\frac{\sqrt{n}(\mu - E_{th})}{\log_2 n}$
Word length n


Empirical standard deviation (σ) and theoretical ($\sigma_{Neininger}$), $n_{exp} = 500$
σ
$\sigma_{Neininger}$
Word length n
Difference between standard deviations, $n_{exp} = 500$
$\Delta\sigma = \sigma - \sigma_{Neininger}$
$(\log_2(n))^{0.50} - 2.58$
$(\log_2(n))^{0.60} - 3.11$
$(\log_2(n))^{0.70} - 3.76$
$(\log_2(n))^{0.80} - 4.55$
$(\log_2(n))^{0.90} - 5.50$
$(\log_2(n))^{1.00} - 6.64$
Word length n