The correlation of a weight with the sum of a series of *k* identically, independently, distributed weights of which it is one

Theory

We assume that the weights are identically, independently distributed according to some distribution for which all moments are defined and that the variance of this distribution is σ_W^2 .

Thus, we require to establish the correlation of $W_{i'}$ with $T=\sum_{i=1}^{k} W_i$, where i' is some integer between 1 and k . Given the assumptions, the following is true

$$Var\left(W_{i'}\right)=\sigma_W^2$$
$$Var\left(T\right)=Var\left(\sum_{i=1}^{k} X_i\right)=k\sigma_W^2$$
$$Cov\left(W_i,W_{i'}\right)=\underset{i=i'}{\sigma_W^2},=\underset{i\neq i'}{0}$$
$$Cov\left(W_{i'},T\right)=\sigma_W^2+\left(k-1\right)0=\sigma_W^2,$$

It thus follows that

$$Corr\left(W_{i'},T\right)=\frac{Cov\left(W_{i'},T\right)}{\sqrt{Var\left(W_{i'}\right)Var\left(T\right)}}=\frac{\sigma_W^2}{\sqrt{\sigma_W^2 k\sigma_W^2}}=\frac{1}{\sqrt{k}}.$$

Simulation

The attached simulations have been carried out in Mathcad® and in Genstat®. Mathcad is much faster but is more limited statistically and as regards plotting. It has been used to do a couple of simulations with larger run numbers. Genstat is slower but has more statistical and plotting capabilities. It has been used to do a number of plots with smaller run numbers but with empirically established confidence intervals. Both sets of plots support that the simple formula above is correct.

To investigate the correlation of a given weight with the sum of weights

Programming language is Mathcad

$ORIGIN := 1$ Origin for vectors and matrices

Theory

$Theory(k) := \frac{1}{\sqrt{k}}$ This is the function that theory suggests applies.

Define various functions

Define function to calculate weight for k trials in a meta-analysis where all trials are of equal size n

$$
Weight(k,n) := \left|
\begin{array}{l}
\text{for } i \in 1..k \\
\quad \left|
\begin{array}{ll}
X \leftarrow rnorm(n,0,1) & \text{Values for } n \text{ patients in trial i} \\
Vmean_i \leftarrow \frac{Var(X)}{n} & \text{Variance of the mean for trial i} \\
Weight_i \leftarrow \frac{1}{Vmean_i} & \text{Weight for trial i}
\end{array}
\right. \\
Weight \qquad \text{Return weights}
\end{array}
\right.
$$

Define function to calculate correlation over m simulations

m is the number of simulations

$$
corrW(k,n,m) := \left|
\begin{array}{l}
\text{for } j \in 1..m \qquad \text{Begin simulation loop} \\
\quad \left|
\begin{array}{ll}
W \leftarrow Weight(k,n) & \text{Calculate } k \text{ weights} \\
T \leftarrow \sum_{i=1}^{k} W_i & \text{Sum the weights} \\
{Y_1}_j \leftarrow W_1 & \text{Assign the first of the } k \text{ weights to value } j \text{ of vector } Y_1 \\
{Y_2}_j \leftarrow T & \text{Assign the total of the weights to value } j \text{ of the vector } Y_2 \text{ for totals of the } k \text{ trials}
\end{array}
\right. \\
rho \leftarrow corr(Y_1, Y_2) \qquad \text{Return correlation over the } m \text{ simulated ocasions}
\end{array}
\right.
$$

Function to carry out 1 to k such correlations

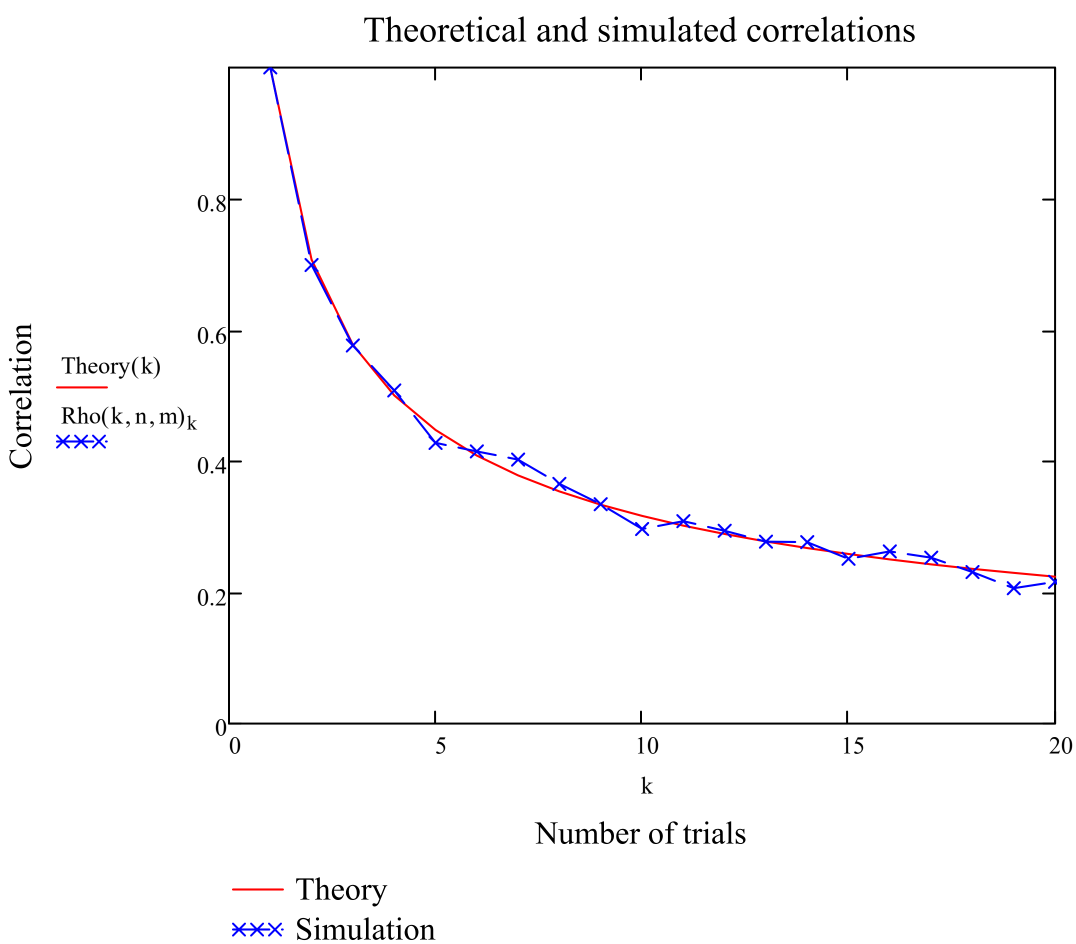
$$\mathrm{Rho}(k,n,m) := \left| \begin{array}{l} \text{for } i \in 1..k \\ \quad \mathrm{Rho}_i \leftarrow \mathrm{corrW}(i,n,m) \\ \mathrm{Rho} \end{array} \right.$$

Plot results

Set parameter values

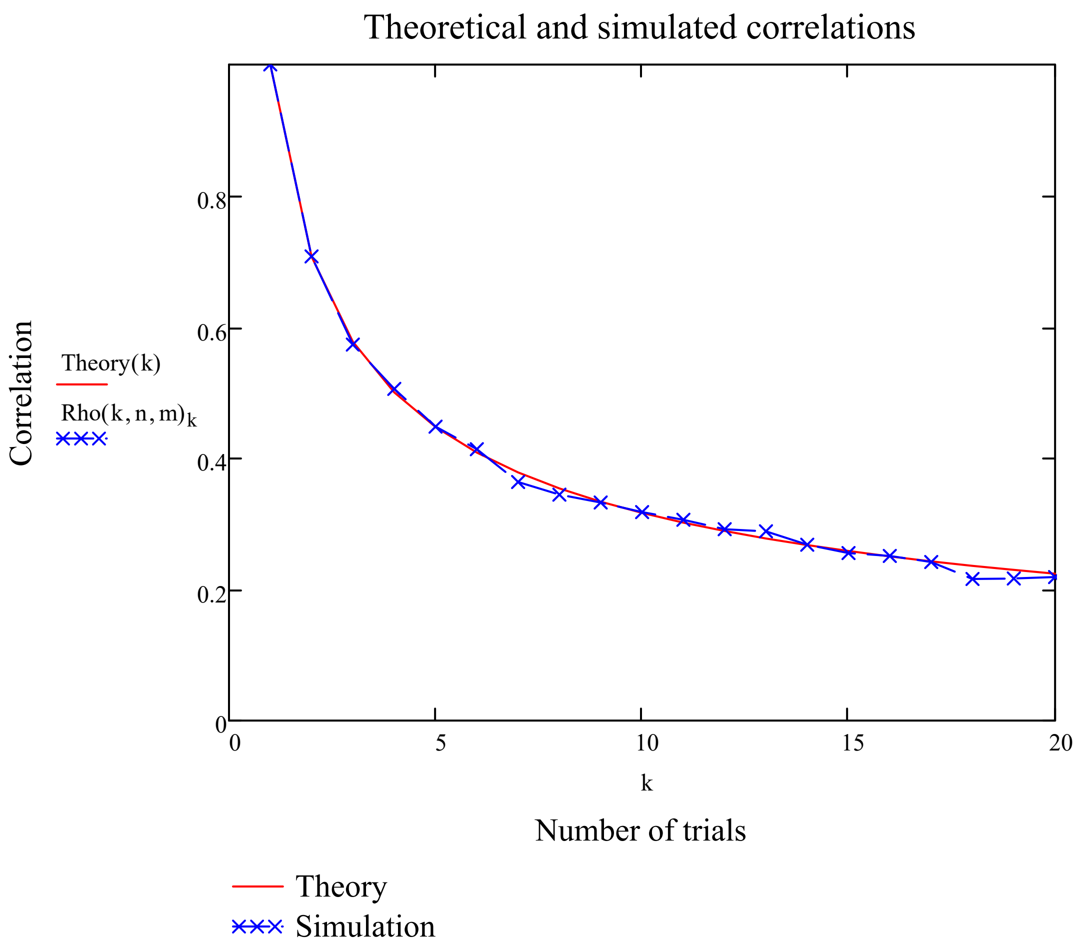
set values for *k* $k := 1..20$ set value for *m* $m := 20000$

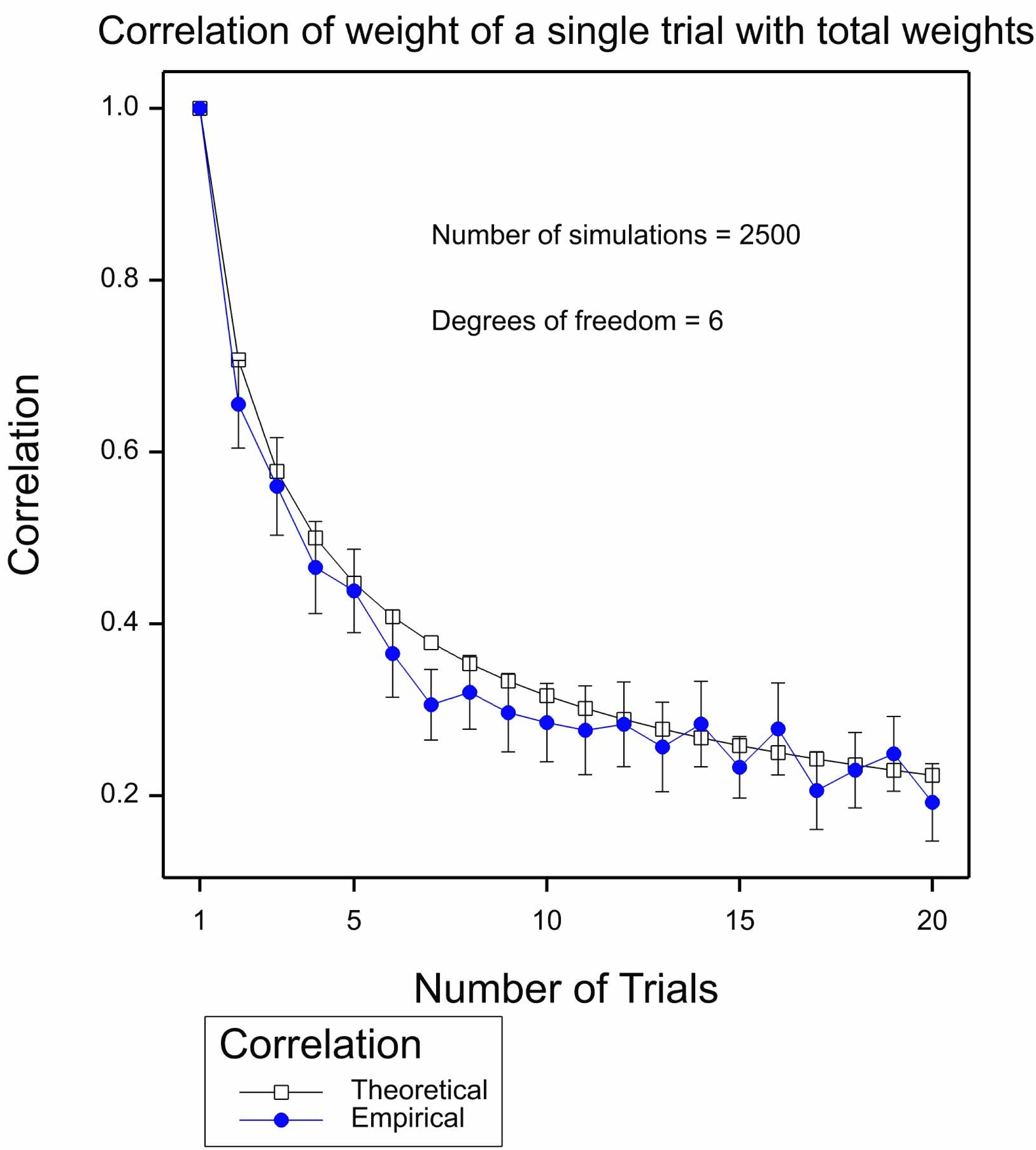
set value for n $n := 10$ degrees of freedom $\nu := n - 1$ $\nu = 9$



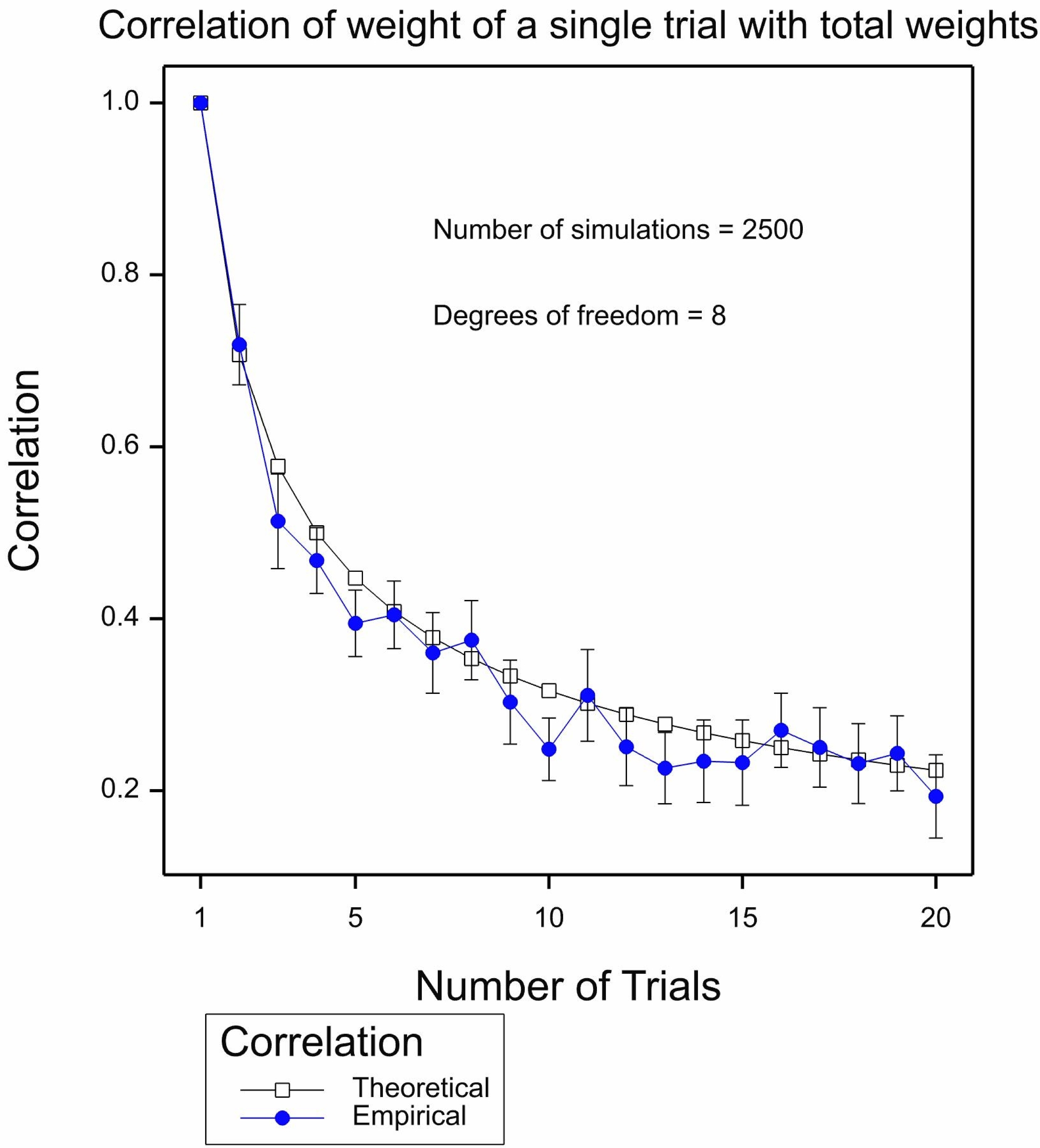
set value for n $\quad n := 20 \quad m := 10000$

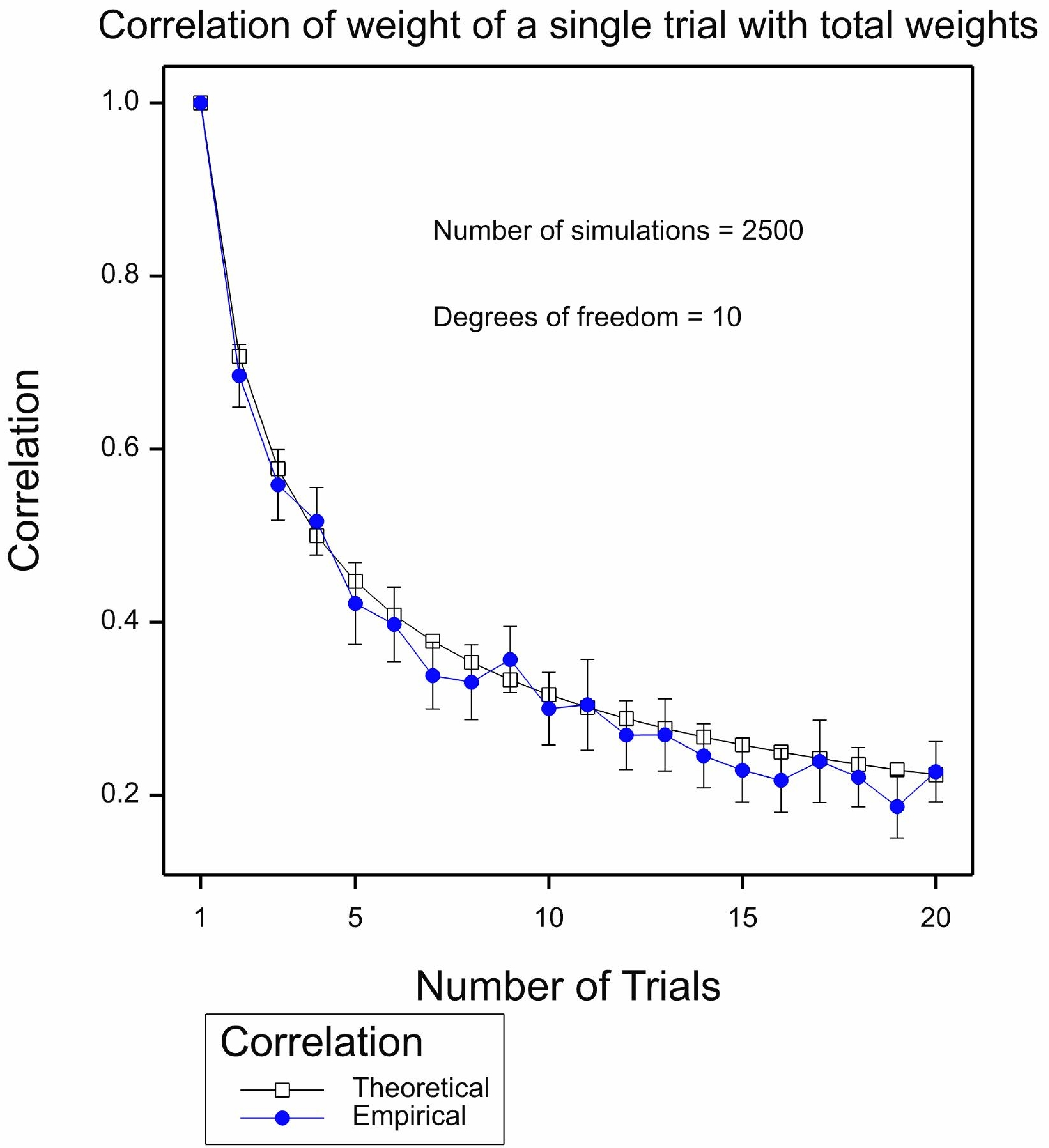
degrees of freedom $\quad \nu := n - 1 \quad \nu = 19$

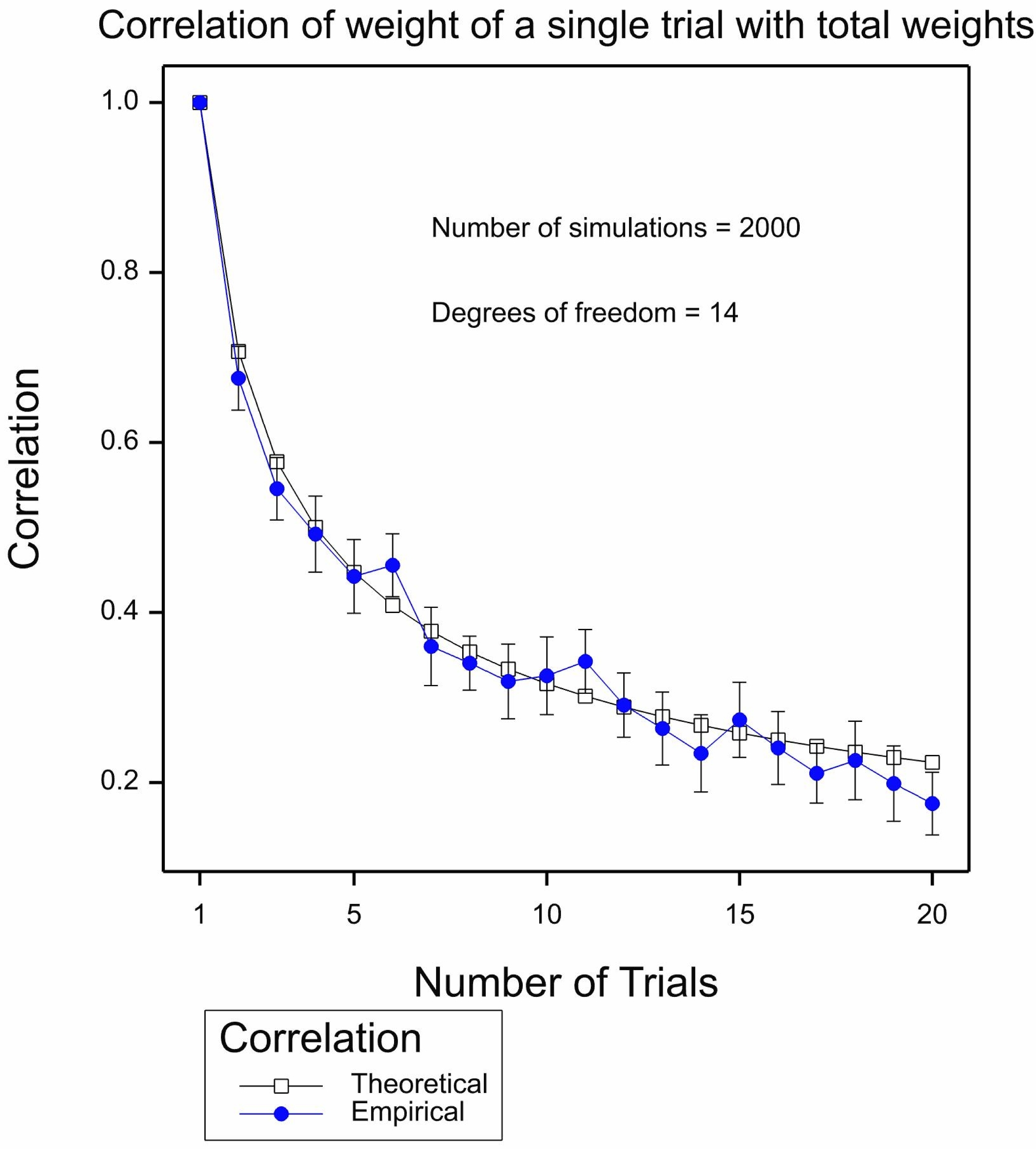


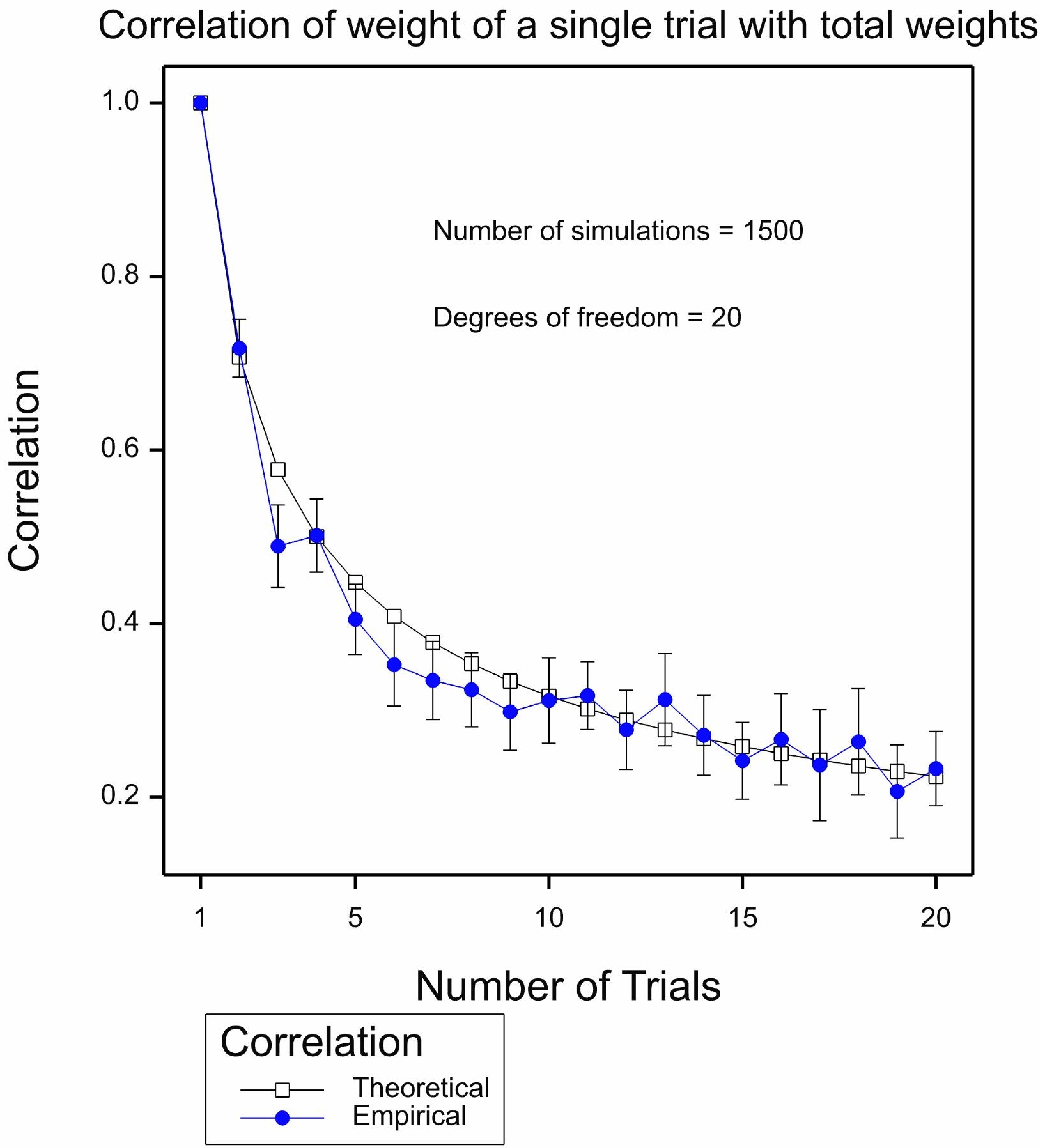


Correlation of weight of a single trial with total weights
Number of simulations = 2500
Degrees of freedom = 6
Correlation
1.0
0.8
0.6
0.4
0.2
1
5
10
15
20
Number of Trials
Correlation
Theoretical
Empirical


Correlation of weight of a single trial with total weights
Number of simulations = 2500
Degrees of freedom = 8
Correlation
1.0
0.8
0.6
0.4
0.2
1
5
10
15
20
Number of Trials
Correlation
Theoretical
Empirical


Correlation of weight of a single trial with total weights
Number of simulations = 2500
Degrees of freedom = 10
Correlation
1.0
0.8
0.6
0.4
0.2
1
5
10
15
20
Number of Trials
Correlation
Theoretical
Empirical


Correlation of weight of a single trial with total weights
Number of simulations = 2000
Degrees of freedom = 14
Correlation
1.0
0.8
0.6
0.4
0.2
1
5
10
15
20
Number of Trials
Correlation
Theoretical
Empirical



Correlation of weight of a single trial with total weights
Number of simulations = 1500
Degrees of freedom = 20
Correlation
1.0
0.8
0.6
0.4
0.2
1
5
10
15
20
Number of Trials
Correlation
Theoretical
Empirical