Confidence intervals (according to Neyman)

Edoardo Milotti

November 28, 2025

The classical construction of frequentist confidence intervals is due to J. Neyman. Neyman's construction formalizes the frequentist way to confidence intervals, starting from the following assumptions:

- The value to be estimated is NOT a random variable, therefore it cannot have a probability density function. The random variables in the frequentist case are the extremes of the confidence interval.
- The pdf of the estimate $\hat{\theta}$ is known for any given value of θ_0 (the *true value of the parameter*).

Here I describe the procedure for a single parameter, as in Neyman's paper.

1. There are two random variables, $\underline{\theta}(D)$ and $\overline{\theta}(D)$, such that

   $$\underline{\theta}(D) < \theta_0 < \overline{\theta}(D)$$

   and

   $$P\left(\underline{\theta}(D) < \theta_0 < \overline{\theta}(D)\right) = \alpha$$

   The quantities $\underline{\theta}(D)$ and $\overline{\theta}(D)$ are random variables because they are functions of other random variables (the data D).

2. There is a precise rule for the construction of $\underline{\theta}(D)$ and $\overline{\theta}(D)$, so that they are known for any given dataset D.

3. The construction starts from the $n+1$-dimensional space $\mathbf{G}$ of spanned by the data vectors and by the parameter values (the dataset D is the result of n measurements): figure 1 shows a conceptual view of this space, where only two data axes are shown.

4. Each given dataset (E in Neyman's notation) is represented by a single point in the hyperplane $\theta = \theta_0$.

5. The same dataset determines a one-dimensional subspace – a straight line parallel to the θ axis that runs through the point corresponding to the dataset. Along this straight line we mark the segment specified by the coordinates $\theta = \underline{\theta}(D)$ and $\theta = \overline{\theta}(D)$

6. We find the set of all datasets such that the corresponding segment and the hyperplane $\theta = \theta_0$ have nonempty intersection. The intersection points (datasets) define a region on this hyperplane (the *acceptance region* $A(\theta_0)$) and mark the segments that do have a nonempty intersection with the hyperplane $\theta = \theta_0$.

7. By construction, each one of these segments is a *covering interval* with confidence level α and corresponds to a given dataset . We see from this construction that

   - If a dataset "belongs" to the acceptance region $A(\theta_0)$, then the corresponding interval covers θ_0
   - If the interval defined by $\underline{\theta}(D)$ and $\overline{\theta}(D)$ covers θ_0, then D "belongs" to the acceptance region $A(\theta_0)$

   In other words, **a dataset belongs to the acceptance region $A(\theta_0)$ if and only if the corresponding interval covers** θ_0. *This means that the probability that a dataset belongs to the acceptance region $A(\theta_0)$ is equal to the probability that the interval defined by $\underline{\theta}(D)$ and $\overline{\theta}(D)$ covers θ_0.*

8. Usually, we do not utilize the full data space, but summarize the information provided by the dataset into an estimator $\hat{\theta} = \hat{\theta}(D)$, and the acceptance region is mapped into a one-dimensional space. Then *the probability that an estimator belongs to the mapped acceptance region $A(\theta_0)$ is equal to the probability that the interval defined by $\underline{\theta}(D)$ and $\overline{\theta}(D)$ covers θ_0.*

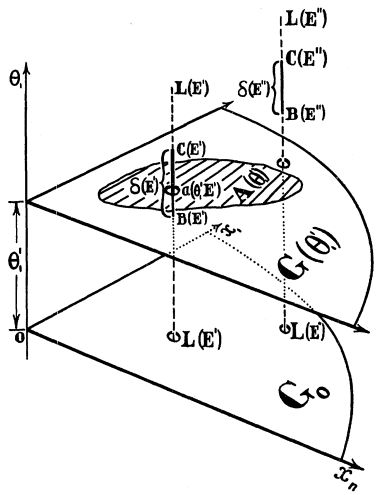

Figure 1: The $n+1$-dimensional space $\mathbf{G}$. Neyman denotes datasets with E instead of D.

A specific example

Now I further illustrate these concepts – and the actual construction of the confidence interval – with n exponentially distributed independent samples. Each sample t_n corresponds to the individual likelihood

$$p(t_n, \theta) = \frac{1}{\tau} \exp\left(-\frac{t_n}{\tau}\right), \tag{1}$$

therefore the global likelihood is

$$L(\{t_n\}, \tau) = \prod_{n=1}^{N} \frac{1}{\tau} \exp\left(-\frac{t_n}{\tau}\right) = \frac{1}{\tau^N} \exp\left(-\frac{\sum_{n=1}^{N} t_n}{\tau}\right), \tag{2}$$

and the log-likelihood is

$$\ln L(\{t_n\}, \tau) = -N \ln \tau - \frac{\sum_{n=1}^{N} t_n}{\tau} \tag{3}$$

We maximize the log-likelihood to find the ML estimate:

$$\frac{d}{d\tau} \ln L(\{t_n\}, \tau) = -\frac{N}{\tau} + \frac{\sum_{n=1}^{N} t_n}{\tau^2} = 0, \tag{4}$$

therefore the ML estimate is the sample average

$$\hat{\tau} = \frac{1}{N} \sum_{n=1}^{N} t_n.$$

It is important to remark that the sample average is an unbiased estimator of the τ parameter, and therefore it is also a sufficient statistic for τ.

The result above shows that the ML estimator $\hat{\tau}$ is the sample mean, and we know that its pdf is a Gamma distribution:

$$p(\hat{\tau}) = \frac{N}{(N-1)!} \frac{(N\hat{\tau})^{N-1}}{\tau^N} e^{-N\hat{\tau}/\tau}. \tag{5}$$

Figure 2 shows the pdf $p(\hat{\tau})$ and a central acceptance interval for a specific true τ.

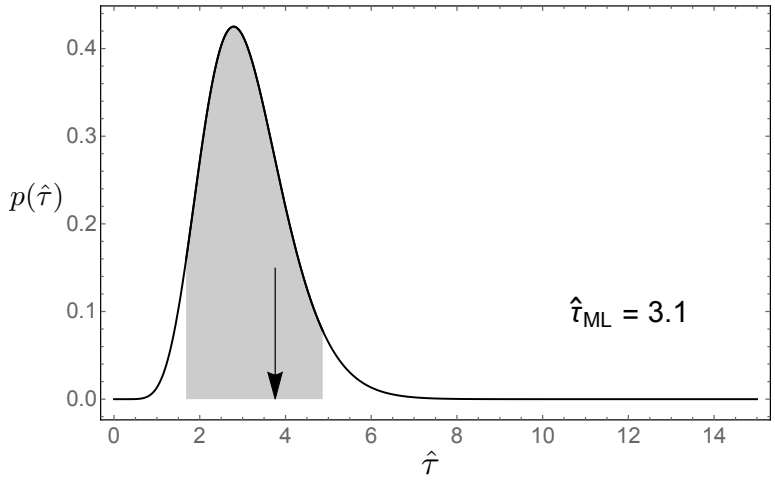

Figure 2: The pdf $p(\hat{\tau})$ and a central interval with 90% confidence level for a specific true value τ. Here $n = 10$, and the true value $\tau = 3.76$ is marked by the arrow.

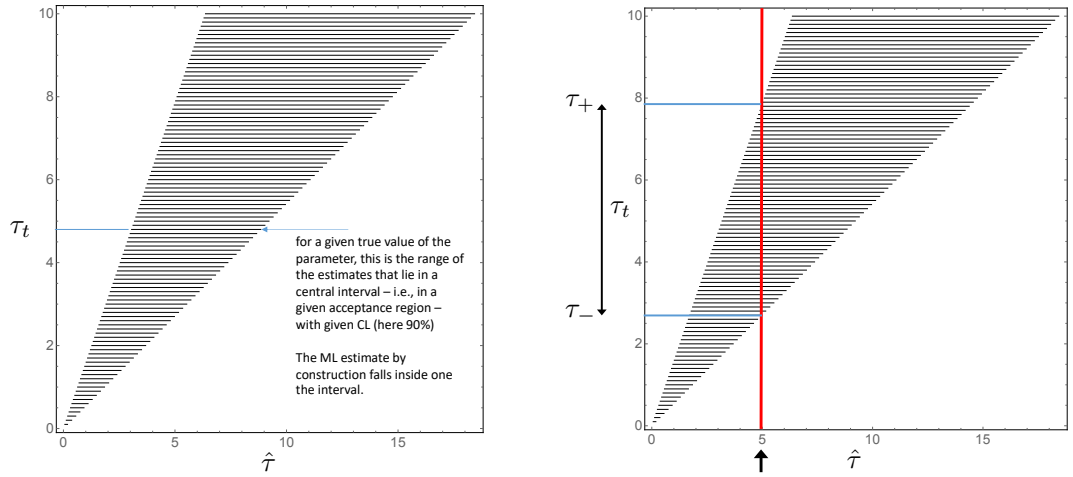

Figure 3: Left panel: the set of intervals that we obtain for different true values τ_t in the example discussed in the main text. Each interval at fixed τ_t represents the acceptance region in the $(\hat{\tau}, \tau_t)$ representation. Right panel: given an estimate of the parameter obtained from actual measurement (shown here by the arrow and the red line) the range of true values that are covered by a central interval with a given confidence level about the estimate (here 90%) has extremes τ_- and τ_+. We know that all the values between τ_- and τ_+ are covered by a central interval about the estimate, with 90% probability. Only the true values between the extremes τ_- and τ_+ are covered according to our predefined rule, therefore the τ_- and τ_+ values are the lower and upper limit of the confidence interval for τ. The collection of all acceptance regions is called the *confidence belt*.

The results framed in the original Neyman construction

We can also go back to the original Neyman construction in the special case of just two measurements (we cannot represent more that a 3D space), i.e., the space **G** is 3-dimensional. Figure 4 shows the 2-dimensional data space with 10000 pairs of (simulated) experiments, figure 5 shows the actual shape of the acceptance region in **G** space, while figure 6 shows the **G**-space representation of a subset of these data with the corresponding central intervals (at 90% confidence level) obtained from the gamma distribution (2) with $n = 2$ (the vertical axis is the τ_t axis). In each figure the red dots marks the pairs of data that belong to the acceptance region.

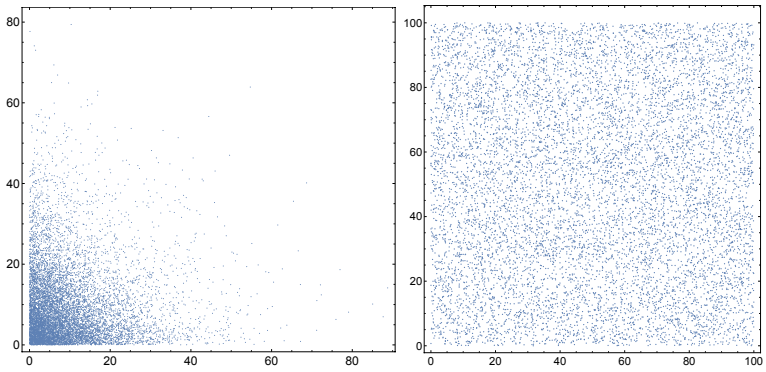

Figure 4: Left panel: simulated distribution of pairs of data with an exponential distribution. Right panel: simulated distribution of data pairs with uniform distribution; this is not a realistic case (actual pairs would be distributed as in the left panel) however this distribution of data pairs is used to fill the plane of data pairs and to better delineate the shape of the acceptance region.

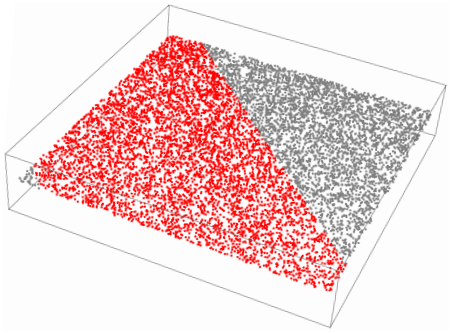

Figure 5: The acceptance region (in the case of the exponential model, as in figure 4) is marked by the red dots.

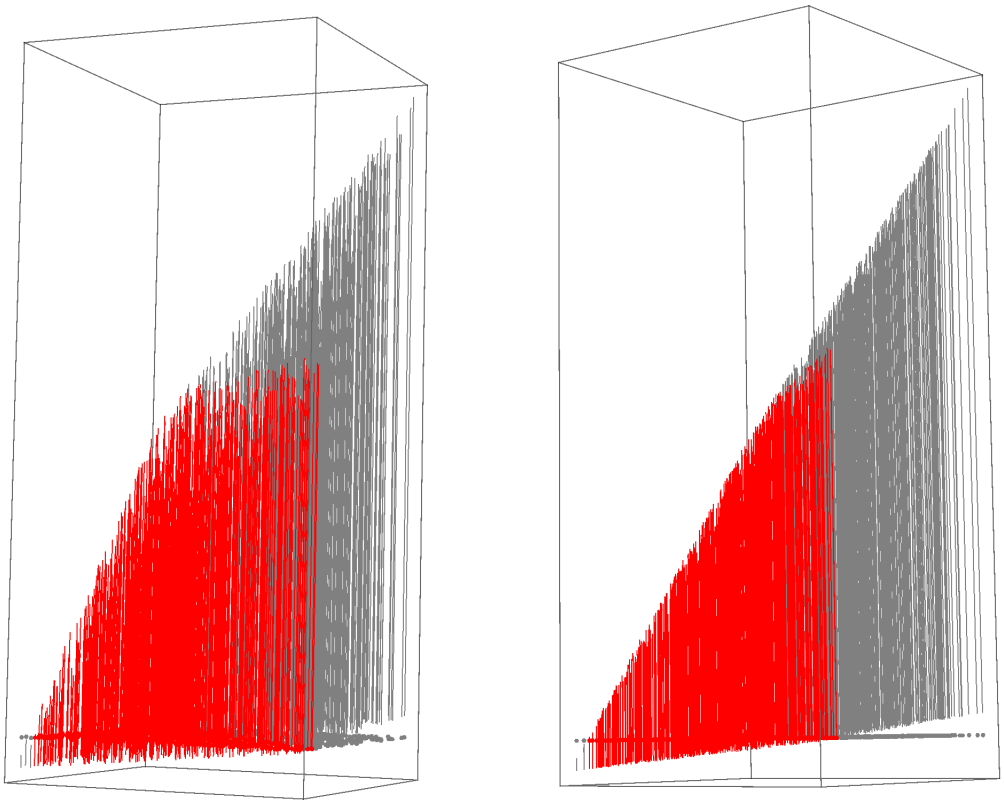

Figure 6: **G**-space representation of the acceptance region (red) with the corresponding central intervals calculated from the gamma distribution (2). The panels show the same data from two different points of view.

The case of the Gaussian pdf

The Neyman construction is simplest in the case of a Gaussian density

$$p(\mu_S, \mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\mu_S - \mu)^2}{2\sigma^2}\right] \tag{6}$$

where μ_S is a sample mean, and where $\mathbf{E}\mu_S = \mu$ (this defines the central line of the confidence belt). In this case it is easy to see that the confidence belt is delimited by straight lines at 45°, and has the simple shape shown in figure 7.

If the underlying distribution of the estimator is Gaussian, it is common to characterize the confidence interval by means of distance from the mean measured in units of standard deviation. In this case we find the confidence levels listed in table 1.

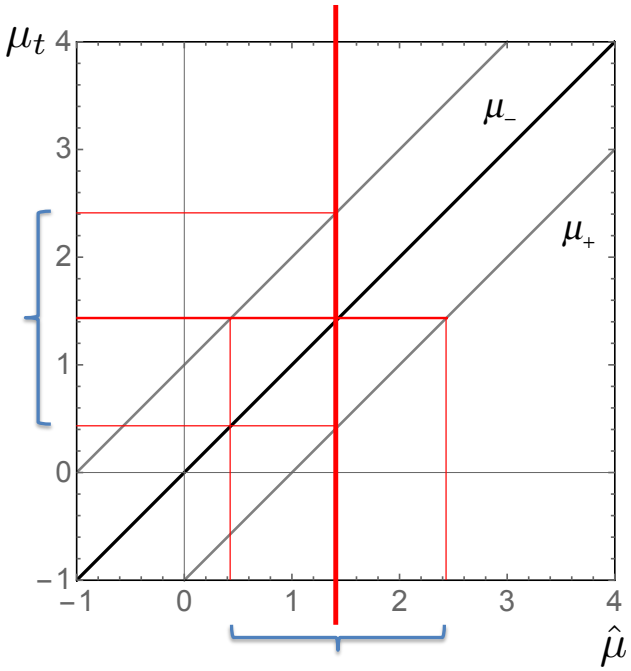

Figure 7: Confidence belt for a Gaussian pdf. Because of the symmetry between the "true value" μ_t and the estimator $\hat{\mu}$ in the Gaussian case, the confidence belt is symmetrical with respect to a 180° flip about the bisector of the first quadrant, and for this reason intervals on the vertical axis are the same as intervals on the horizontal axis (acceptance regions have the same extent as the confidence intervals).

The Gaussian distribution is often used as an approximation of more complex cases, but this does not always work. Figure 8 (right panel) shows the lower and upper values of the central 95% confidence intervals for the sample mean of the exponential distribution, compared with the corresponding intervals obtained with the Gaussian approximation of the gamma distribution. Notice that the Gaussian confidence interval approximates quite well the interval from the gamma distribution, however for small n the lower bound can extend to negative – and unphysical – values.

Table 1: Confidence level for a Gaussian distribution, and intervals at $\pm k\sigma$ about the mean, with 10 digits accuracy.

k	confidence level
1	0.6826894921
2	0.9544997361
3	0.9973002039
4	0.9999366575
5	0.9999994267
6	0.9999999980

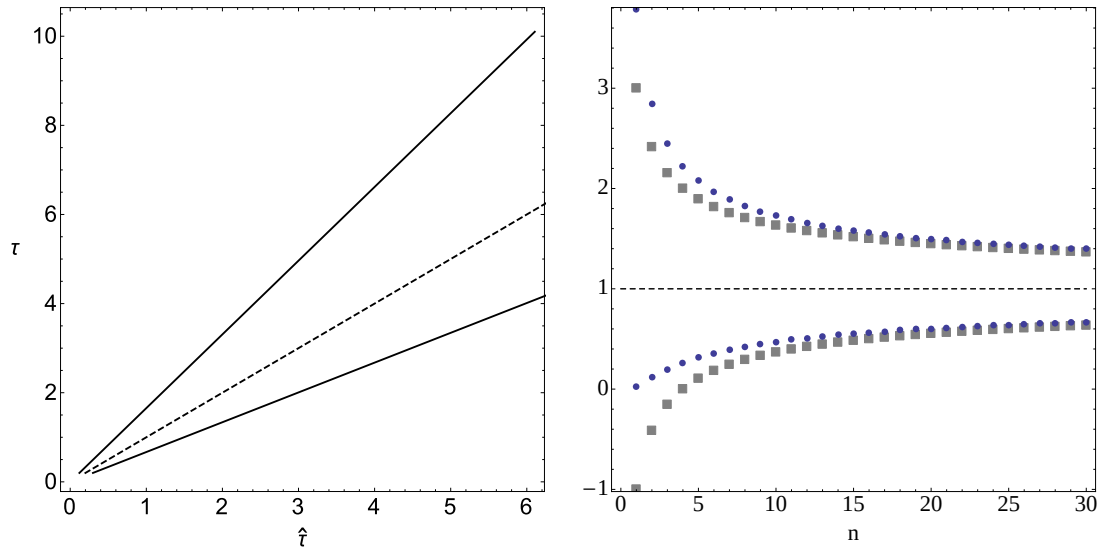

Figure 8: Left panel: confidence belt for the mean of exponentially distributed data (τ vs. $\hat{\tau}$). The solid lines show the lower and upper limits of the central interval with a confidence level $C = 0.95$, while the dashed line shows the position of the "true" mean τ. Right panel: plot of lower and upper values of the central 95% confidence intervals for the sample mean of sets of exponentially distributed data vs. the sample size n (circles). Here $C = 0.95$. The corresponding values for the Gaussian approximation (2σ intervals) are also shown (squares).

Confidence intervals for a more complex estimator: the correlation coefficient

Even in the simple case of a bivariate Gaussian distribution, the pdf of the sample estimator of the correlation coefficient – for a given estimate r_S obtained from data – is quite complex, and definitely non-Gaussian[1]:

$$p(r, r_S) = \frac{1}{\pi}(n-2)(1-r^2)^{(n-4)/2}(1-r_S^2)^{(n-1)/2}\int_0^\infty \frac{d\beta}{(\cosh\beta - r\, r_S)^{n-1}} \tag{7}$$

However in cases such as this (and more complex) we can determine the confidence interval from an MC simulation that yields an empirical pdf for the parameter. We use this empirical distribution much like the gamma distribution used to determine the confidence belt. Figure 9 shows 10000 points generated according to a bivariate Gaussian distribution: the distributions of the x and y coordinates are shown in figure 10. The correlation coefficient of each data set can be estimated from the formula

$$r_S = \frac{\mathrm{cov}_S(x,y)}{\sigma_{x,S}\sigma_{y,S}} = \frac{\sum_{k=1}^n (x_k - \mu_{x,S})(y_k - \mu_{y,S})}{\sqrt{\sum_{i=1}^n (x_i - \mu_{x,S})^2 \sum_{j=1}^n (y_j - \mu_{y,S})^2}} \tag{8}$$

and this is shown in figure 11. The MC distribution allows the determination of the limits of the central confidence interval, as shown in figure 12.

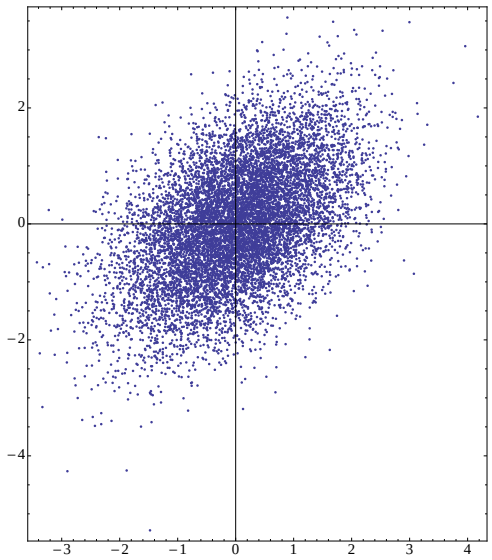

Figure 9: 10000 points generated according to a bivariate Gaussian distribution with $r = 0.5$.

[1] see, e.g., http://mathworld.wolfram.com/CorrelationCoefficientBivariateNormalDistribution.html

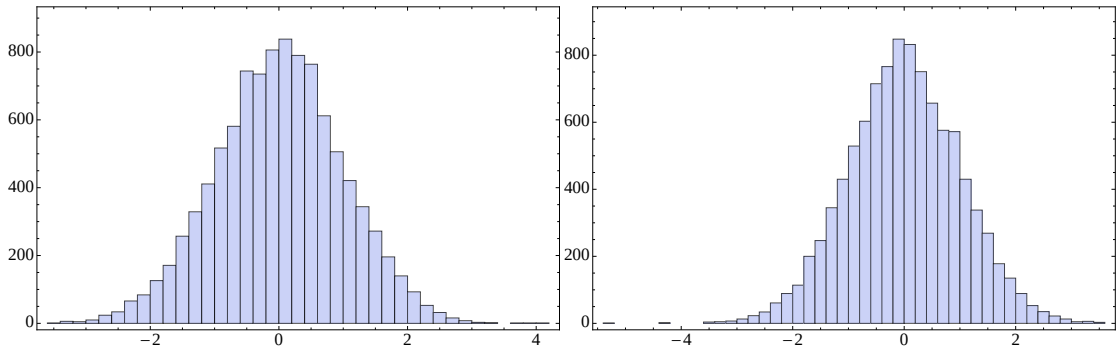

Figure 10: Individual histograms of the x and y coordinates of the points in figure 9.

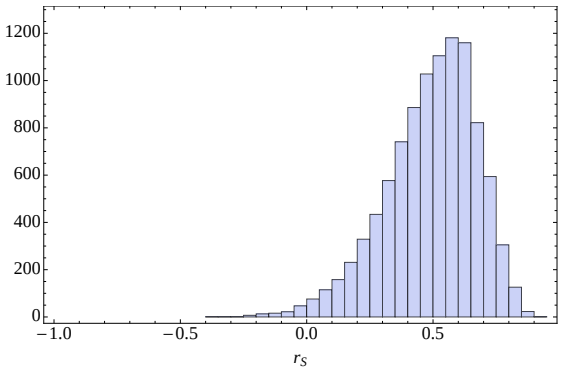

Figure 11: Histogram of the sample correlation coefficient for 10000 datasets of 20 samples each.

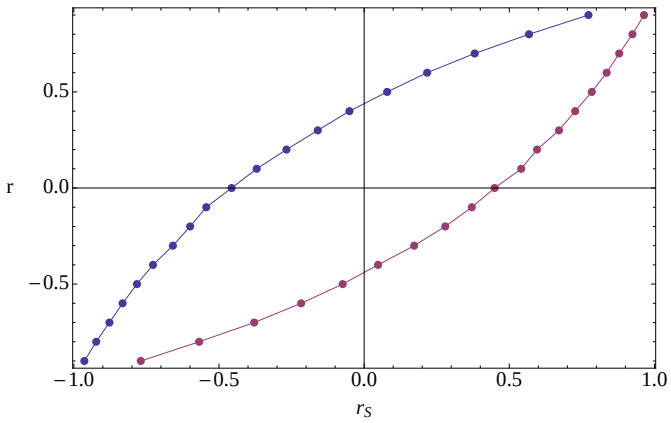

Figure 12: 95% confidence belt for the correlation coefficient of 20 samples, from the MC calculation.